THE UNIVERSITY *of* EDINBURGH

Raman Spectroscopy as a tool to improve Enhanced Biological Phosphorus Removal

Helen A. Cope

A thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

School of Engineering

2016

Declaration

I declare that this thesis has been composed by myself and that the work contained in it is my own, except where stated otherwise. Furthermore, this work has not been submitted for any other degree or professional qualification.

Helen Cope

,

Abstract

Enhanced Biological Phosphorus Removal (EBPR) is an established process in wastewater treatment that uses bacteria to reduce phosphorus levels below regulatory discharge limits. Recently, in light of growing political concern over phosphorus sustainability, EBPR has also been recognised as a platform from which phosphorus may be recovered and recycled onto land as fertiliser. Operating EBPR to optimise performance and efficiency is therefore extremely important, but remains a challenge due to poor understanding of these bacterial ecosystems. Raman spectroscopy is a non-invasive, label-free, culture-independent technique capable of analysing live, single cells. Despite its advantages, Raman spectroscopy has been applied to study EBPR bacteria in just a handful of studies and thus has a low profile in this field of research. More work is required to investigate potential areas of application for Raman spectroscopy in EBPR research. The principal thesis presented here is that Raman spectroscopy could be used as a tool to improve EBPR.

The Raman spectra used for this investigation were acquired from individual EBPR bacteria dried onto a calcium fluoride substrate. The bacterial samples were collected from three different sources, namely lab-scale sequencing batch reactors located in Edinburgh (University of Edinburgh, UK) and Boston (Northeastern University, USA), and a full-scale EBPR plant in Slough (Thames Water, UK). Using these spectra, some potential applications and limitations of Raman spectroscopy for improving EBPR were explored. In this foundation work, a particular emphasis on spectral analysis methods was kept in light of the benefits of automating analysis as well as the need for standardisation to be able to compare results between different studies and groups.

Nine methods were compared for baselining Raman spectra of individual EBPR bacteria. From these, the “small-window moving average” (SWiMA) method was determined to be the best baselining technique for our purposes at the current time. In agreement with earlier studies, the Raman spectroscopic signatures of three key EBPR metabolites – polyphosphate, polyhydroxyalkanoate (PHA) and glycogen – were shown to be clearly identifiable in individual EBPR bacteria when present. The Raman shifts of characteristic spectral bands arising from polyphosphate were shown to vary significantly between samples and the implications of this were discussed. Examples of how the Raman spectra of individual bacteria can be modelled with multivariate tools to open up new areas for research were given. MCR modelling was demonstrated to offer a novel way to normalise the Raman spectra of individual EBPR bacteria prior to quantitative analysis. With the instrumental set-

up in this work, the limit of detection (LOD) of aqueous polyphosphate samples was estimated to be approximately 0.08 M and 0.02 M for 10 second and 200 second acquisitions respectively. Future work is required to research ways in which a more comparable form of polyphosphate ‘standard’ might be prepared so that direct correlation can be drawn between measurements made on such a standard and measurements made in bacterial cells.

Overall, several applications and challenges of Raman spectroscopy for the investigation of EBPR bacteria are presented in this work together with recommendation for how to process the spectral data. The conclusions drawn from this work indicate that Raman spectroscopy could be used as a tool to improve EBPR but further work is required to refine and apply these methods.

Here I raise my Ebenezer

Hither by Thy help I've come

— Robert Robinson —

Acknowledgements

This work could not have been completed without the help and support of many people.

My supervisors Alistair Elfick and Andy Downes and the School of Engineering gave me the opportunity, coaching and financial support to pursue this PhD research right through to its successful completion. Thank you for always being actively on my side.

Ioan Notingher (University of Nottingham) and Dimitri Mignard took the time to examine me on my work and thus enabled me to cross the finishing line. Thank you for your time, your interest and all your comments.

Steve Gourlay, Bill Leslie, Kevin Tierney, Douglas Carmichael and Paul Aitken provided vital mechanical and electrical technical support to get my experimental rig up and running. Thank you for all your advice, patience and hard work.

April Gu, Annalisa Onnis-Hayden and Yueyun Li at Northeastern University, Boston (MA) generously gave me the opportunity to work with them for four months to learn all things EBPR. Thank you for welcoming me and being excellent hosts. I look forward to continuing work with you in future.

Andrea Semião and Jana Schwarz-Linek provided regular access to instruments necessary for monitoring which were not available in our own lab. Paula Lindgren (University of Glasgow) gave me both the opportunity and support to acquire SEM images. Tony Corcoran provided AFM images. Andrew Free gave vital biological advice and worked with Helen Williamson to analyse selected samples for DNA, the results of which we hope to use in future. Rosanna Kleemann (University of Surrey) periodically sent samples of EBPR activated sludge with which to inoculate my reactors. She also kindly showed me around the Thames Water wastewater treatment plant at Slough. Vlastimil Srsen was a tireless source of help, encouragement and support around the lab.

Alex Cao (University of Michigan), Paul Stoddart (Swinburne University of Technology) and Georg Schulze (University of British Colombia) provided MATLAB scripts of their spectral baselining methods via personal communication. Thank you for sharing your work and in doing so enable me to broaden the suite of methods I was able to trial on my spectra.

Rasmus Bro (University of Copenhagen) and Gordon Reid (Scottish Water) provided very generous and fruitful discussions despite having no obligation toward me. Thank you both

for inviting me to visit and discuss my work and questions in person. Thank you Rasmus for hosting me in Copenhagen and giving me the opportunity to present to your group.

Funding for this PhD was largely provided by the Engineering and Physical Sciences Research Council. A £4,750 J M Lessells Postgraduate Scholarship in Engineering from the Royal Society of Edinburgh enabled me to conduct research in Boston for four months. Several travel scholarships from the Society for General Microbiology and the Royal Society of Chemistry have enabled me to attend conferences in Dublin, Montpellier and Sydney. For all this support I am deeply grateful.

My family and personal friends have been cheering me on from the beginning. Thank you for always believing in me and for carrying me in so many ways. Liane Benning, you have chased me to do a PhD for a very long time. Thank you for some very practical advice along the way. Thank you Domi for many years of encouragement, advice and adventures. Thank you Scott and Louise for walking so closely with me these last few years. Thank you Row and Jon, and Paul and Pam – you know the very practical ways in which you have helped and encouraged me through some of the harder days. Thank you Ares for sustaining me with your friendship and the contents of your food drawer. Thank you Wesam for your encouragement and for submitting this thesis in my absence when I spent the last week before the deadline writing furiously in Northern Cyprus. Thank you İme, Gülser and Mehmet for hosting me and carrying me through the last days. I will not forget your love, generosity and nourishment – including *nar* and *zeytin*. *Unutma beni* – I will be back! Thank you mum and dad for giving me a firm foundation and for always supporting me in my ventures. Thank you for always being there for me.

This success is not mine alone. I share it with all of you.

Abbreviations

AKBS	adaptive knots B-spline
AMI	amide
ALS	asymmetric least squares
AMM	adaptive minmax
ANN	artificial neural networks
APLS	adaptive-weight penalised least squares
arb.	Arbitrary
ATCC	American Type Culture Collection
ATR-FTIR	attenuated total reflectance Fourier transform infrared spectroscopy
BE	beam expander
BKG	background
CCD	charge-coupled device
COD	chemical oxygen demand
cPAO	candidate polyphosphate accumulating organism
DAPI	4’,6-diamidino-2-phenylindole
DAQ	data acquisition
deg	polynomial degree
DG	diffraction grating
DO	dissolved oxygen
EBPR	enhanced biological phosphorus removal
EMSC	extended multiplicative scatter correction
FISH	fluorescence in situ hybridisation
FS	fluorescence to signal (ratio)
FT	Fourier transform

FWHM	full width at half maximum
GAO	glycogen accumulating organism
GLY	glycogen
HNF	holographic notch filter
HRT	hydraulic retention time
IR	infrared
iter	number of iterations
L	lens
LED	light emitting diode lamp
LM	local medians
LOWESS	locally weighted scatterplot smoother
LV	Latent variable
M	mirror
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight
MATLAB	matrix laboratory (software)
MCR	multiplicative curve resolution
ML	mixed liquor
MLVSS	mixed liquor volatile suspended solids
ModPoly	modified polynomial
MSC	multiplicative scatter correction
n	number
ND	neutral density filter
NIR	near-infrared
NMR	nuclear magnetic resonance
OB	microscope objective
OHO	ordinary heterotrophic organism

OP	orthophosphate
OSC	orthogonal signal correction
P	phosphorus
PAO	polyphosphate accumulating organism
PF	plasma filter
PHA	polyhydroxyalkanoate
PHB	polyhydroxybutyrate
PHE	phenylalanine
PHV	polyhydroxyvalerate
PLS	partial least squares
PLS-DA	partial least squares discriminant analysis
Poly	simple polynomial
PP	polyphosphate
QIA	quantitative image analysis
R^2	correlation coefficient
RB	rolling ball
RBE	robust baseline estimation
REFS.	references
roi	region of interest
rpm	revolutions per minute
S	slit
SBR	sequencing batch reactor
SD	standard deviation
S/N ratio	signal to noise ratio
SNV	standard normal variate
SPG45	sodium phosphate glass type 45

SRT	solids retention time
SWiMA	small-window moving average
TM	triangular mirror
tol	tolerance of difference between iterations
TSS	total suspended solids
VSS	volatile suspended solids
WT	wavelets transform
WWTP	wastewater treatment plant

Table of Contents

ABSTRACT III

ACKNOWLEDGEMENTS VII

ABBREVIATIONS IX

TABLE OF CONTENTS XIII

LIST OF FIGURES XVII

LIST OF TABLES XXIII

CHAPTER 1 INTRODUCTION 1

1.1 PHOSPHORUS IN CONTEXT 3
1.2 OVERALL AIM AND RESEARCH QUESTIONS 6
1.3 THESIS OVERVIEW 7

CHAPTER 2 LITERATURE REVIEW 9

2.1 ENHANCED BIOLOGICAL PHOSPHORUS REMOVAL 11
2.2 RAMAN SPECTRA OF BACTERIA 15
2.3 RAMAN SPECTRA APPLIED TO EBPR 18

CHAPTER 3 METHODS I: DESIGN, BUILD AND VALIDATION OF SEQUENCING BATCH REACTORS FOR ENHANCED BIOLOGICAL PHOSPHORUS REMOVAL 21

3.1 REQUIREMENT FOR SEQUENCING BATCH REACTORS 23
3.2 DESIGN AND BUILD OF SEQUENCING BATCH REACTORS 25
3.2.1 Specifications *25*
3.2.2 Design and Build *26*
3.3 METHODS FOR MONITORING 32
3.3.1 Determination of Orthophosphate *32*
3.3.2 Determination of Chemical Oxygen Demand (COD) *33*
3.3.3 Determination of Suspended Solids *33*
3.3.4 Measurement of Dissolved Oxygen (DO), pH and Temperature *34*
3.3.5 Neisser Staining for Microscopy *34*
3.3.6 Scanning Electron Microscopy (SEM) *34*
3.4 VALIDATION OF SEQUENCING BATCH REACTORS 35
3.4.1 Operating Conditions *35*
3.4.2 Performance *44*
3.4.3 Second Run of the SBRs *46*
3.4.4 Summary of Modifications made during SBR Operation *51*

3.5 FUTURE IMPROVEMENTS FOR SEQUENCING BATCH REACTORS 52

CHAPTER 4 METHODS II: RAMAN SPECTROSCOPY OF INDIVIDUAL BACTERIA 55

4.1 INTRODUCTION TO RAMAN SPECTROSCOPY 57

4.1.1 Molecular Vibration 57
4.1.2 The Raman Effect 59
4.1.3 The Raman Spectrum 62
4.1.4 Confocal Raman Micro-Spectroscopy (Raman Microscopy) 65

4.2 ACQUISITION OF RAMAN SPECTRA 66

4.2.1 Instrumentation 66
4.2.2 Spectral Resolution 68
4.2.3 Spatial Resolution 70
4.2.4 Experimental Estimation of Raman Sampling Volume 71
4.2.5 Instrument Alignment 78
4.2.6 Acquisition Parameters 78
4.2.7 Sample Preparation 80

4.3 SPECTRAL PREPROCESSING 81

4.3.1 Cosmic Spike Removal 81
4.3.2 Background Removal and Baselining 81

4.4 OVERVIEW OF SPECTRAL ANALYSIS 82

4.4.1 General Methods 82
4.4.2 Multivariate Methods 82

CHAPTER 5 BASELINING RAMAN SPECTRA OF INDIVIDUAL EBPR BACTERIA 85

5.1 INTRODUCTION 87

5.1.1 Character of EBPR Bacteria Spectra 88
5.1.2 Approaches to Baselining 90
5.1.3 Selection of Methods for Comparison 94

5.2 METHODOLOGY 95

5.2.1 Spectroscopic Measurements 95
5.2.2 Data Sets 95
5.2.3 Preprocessing Methods 95
5.2.4 Evaluation of Baselining Methods 103

5.3 RESULTS AND DISCUSSION 104

5.3.1 Appearance of Corrected Spectra 104
5.3.2 Potential Effect on Classification 109
5.3.3 Potential Effect on Quantification of Polyphosphate 116
5.3.4 Other Considerations 118
5.3.5 Summary and Conclusions 119

CHAPTER 6 RAMAN SPECTRA OF INDIVIDUAL EBPR BACTERIA: A QUALITATIVE EXAMINATION 121

6.1 INTRODUCTION 123

6.2 MATERIALS AND METHODS 123

6.2.1 Spectroscopic Measurements 123
6.2.2 Data Sets 123

6.2.3 Analysis *125*
6.3 IDENTIFICATION OF KEY EBPR METABOLITES 126
6.3.1 Characteristic Spectra of Bacteria, Polyphosphate, PHAs and Glycogen *126*
6.3.2 Variation in Peak Position of Key Metabolites between Samples *136*
6.3.3 Identification of Key Metabolites with MCR *139*
6.3.4 Morphology of Individual Cells *145*
6.4 AUTOMATED IDENTIFICATION OF CANDIDATE PAOS 147
6.4.1 Identification of cPAOs Based on Polyphosphate Peak Height *148*
6.4.2 Identification of cPAOs Based on PLS-DA *154*
6.4.3 Comparison of Methods for Automated Identification of cPAOs *159*
6.5 SUMMARY AND CONCLUSIONS 160

CHAPTER 7 QUANTIFICATION OF POLYPHOSPHATE IN EBPR BACTERIA 163
7.1 INTRODUCTION 165
7.2 METHODOLOGY 171
7.2.1 Normalisation Methods for Quantification of Polyphosphate in EBPR Bacteria *171*
7.2.2 Estimation of Limit of Detection and Limit of Identification *174*
7.2.3 Spectroscopic Measurements *174*
7.2.4 Data Sets *174*
7.2.5 Spectral Analysis *175*
7.3 PERFORMANCE OF NORMALISATION METHODS 177
7.4 ESTIMATION OF LOD AND LOI OF POLYPHOSPHATE 188
7.4.1 LOD and LOI of Aqueous Polyphosphate Solutions *188*
7.4.2 LOD of Polyphosphate in EBPR Bacteria *191*
7.5 SUMMARY AND CONCLUSIONS 194

CHAPTER 8 SUMMARY AND CONCLUSIONS 195

APPENDIX 1 SBR RIG COMPONENTS 203

APPENDIX 2 COMPARISON OF PARAMETERS FOR SELECTED BASELINING METHODS 207

REFERENCES 213

List of Figures

Figure 1-1 Geographical distribution of global phosphate rock reserves in 2015. 3

Figure 1-2 The phosphorus ‘chain’. 4

Figure 2-1 Principal biochemical transformations occurring in PAOs during EBPR..... 12

Figure 2-2 Concentration profiles of key compounds during the anaerobic and aerobic phases of EBPR. 13

Figure 3-1 Schematic diagram depicting the layout of the Sequencing Batch Reactors... 27

Figure 3-2 Photographs depicting the SBR layout 28

Figure 3-3 Schematic diagram of the control and monitoring interface of the Sequencing Batch Reactors..... 29

Figure 3-4 Components operational during each phase of the Sequencing Batch Reactor cycle 30

Figure 3-5 EBPR lanes at Slough Sewage Treatment Works, UK 36

Figure 3-6 TSS and VSS levels in SBR-A and SBR-B during the validation phase. 39

Figure 3-7 Temperature, pH and Dissolved Oxygen for SBR-A and SBR-B during the validation phase. 40

Figure 3-8 Paired plots of mean pH and mean dissolved oxygen (DO) measured during the aerobic phase for SBR-A and SBR-B 42

Figure 3-9 Micrographs of mixed liquor from SBR-A and SBR-B 44

Figure 3-10 PO_4-P removal performance of SBR-A and SBR-B measured during individual anaerobic-aerobic cycles during the validation phase. 46

Figure 3-11 TSS and VSS levels in SBR-C and SBR-D during the second run of the SBRs. 47

Figure 3-12 Temperature, pH and Dissolved Oxygen for SBR-C and SBR-D during the second run of the SBRs. 48

Figure 3-13 Micrographs of mixed liquor from SBR-C and SBR-D 49

Figure 3-14 SEM micrographs of mixed liquor sampled from SBR-D 49

Figure 3-15 PO_4-P removal performance of SBR-C and SBR-D measured during individual anaerobic-aerobic cycles during the second run of the SBRs. 50

Figure 4-1 Change in polarisability and hence Raman ‘activity’ with the molecular vibrational modes of a linear triatomic molecule. 58

Figure 4-2 Schematic representation of Rayleigh (elastic) and Raman (inelastic) scattering of light.....60

Figure 4-3 Diagrammatic representation of molecular energy levels and transitions corresponding to Infrared, Rayleigh and Raman scattering.....61

Figure 4-4 Example Raman spectrum of the tetrahedral molecule carbon tetrachloride (CCl_4).....62

Figure 4-5 Raman shifts of some common bond vibrations.....64

Figure 4-6 The Renishaw InVia Raman microscope used in this work 67

Figure 4-7 Change in position of a small object as a function of time when illuminated by a hot light source (100 W halogen lamp) versus a cold light source (LED array lamp).....68

Figure 4-8 Raman spectra of neon for experimental determination of spectral resolution. 69

Figure 4-9 Sample curve for 'GaussAmp' curve fitting function in Origin, diagrammatically showing parameters relevant to both 'GaussAmp' and 'InvsPoly' fits. 72

Figure 4-10 Experimental estimation of laser spot size in the lateral plane using 5 µm polystyrene spheres deposited on a quartz substrate.....73

Figure 4-11 Experimental estimation of the axial range of the focussed laser spot using different materials.....76

Figure 4-12 Raman line maps through two example bacterial cells.....80

Figure 4-13 Schematic diagram and image of a custom made aluminium slide for presentation of aqueous samples.....81

Figure 5-1 Raman spectra of four individual EBPR bacteria offset for clarity (a). The same spectra are shown individually together with a spectrum of the pure EBPR storage compound they contain – polyphosphate (b), PHB (a type of PHA) (c), glycogen (d), and no visible storage compound (e). The instrument dark spectrum and calcium fluoride substrate spectrum are shown in (f). 89

Figure 5-2 Framework and overview of some approaches to baseline correction. 91

Figure 5-3 (Left) Diagrammatic representation of the sampling volume of a Raman instrument laser encompassing both the bacterium and the substrate. (Right) Depiction of the orthogonal vectors of the substrate **S** and bacterium **B** spectra, and the combined substrate and bacterium spectrum **C**.....97

Figure 5-4 Polynomial curve fitting as it is applied to rhodamine 6G. 99

Figure 5-5 One background subtraction and nine baselining methods demonstrated on a single EBPR bacterium Raman spectrum.....106

Figure 5-6 Five different baselining methods demonstrated on four different background-corrected EBPR bacterial Raman spectra..... 108

Figure 5-7 Loadings for 4-component MCR models built on a test set of 158 Raman spectra of individual EBPR bacteria. The Raman spectra were preprocessed seven different ways..... 110

Figure 5-8 Comparison of MCR model components (in blue) and pure compound spectra (in red). 111

Figure 5-9 32 Raman spectra of a calcium fluoride substrate and the same spectra normalised to unit area. 112

Figure 5-10 Depiction of the cross-validation specificity and sensitivity for classification of 158 Raman spectra of individual EBPR bacteria based on seven different PLS-DA models..... 115

Figure 6-1 Pictorial representations and chemical structures of three key EBPR metabolites. 127

Figure 6-2 Raman spectra of a dried cell-pellet smear of Pseudomonas putida and of an individual EBPR bacterium. 128

Figure 6-3 Raman spectra of sodium phosphate glass type 45 (SPG45) and polyphosphate in an EBPR bacterium. 130

Figure 6-4 Linear structure of polyphosphate with PO_2^- moieties circled in red. 130

Figure 6-5 Raman spectra of the pure compounds polyhydroxybutyrate (PHB) and polyhydroxybutyrate-co-(12%)-polyhydroxyvalerate (PHB-co-(12%)-PHV) and of an EBPR bacterium containing PHA. 132

Figure 6-6 Raman spectra of pure glycogen and glycogen in an EBPR bacterium. 135

Figure 6-7 Box plots showing statistics for peak position of polyphosphate bands 1 (at ~697 cm^{-1}) and 2 (at ~1169 cm^{-1}) in the subsets of bacteria indicated. 138

Figure 6-8 Loadings for the 4-component MCR model built as described in Section 6.2.3. 140

Figure 6-9 Scores plotted for a 4-component MCR model built as described in Section 6.2.3 on a calibration dataset (top), and applied subsequently to the EDI100, NE100 and SLO100 datasets..... 142

Figure 6-10 Comparison for each data set of original (summed) spectra and the (summed) spectra reconstructed from the MCR model..... 144

Figure 6-11 Summed residuals from the MCR model applied to each of the four data sets 145

Figure 6-12 Example micrographs (magnification ×1000) of individual EBPR bacteria dried on a CaF_2 substrate and subsequently analysed with Raman spectroscopy. 146

Figure 6-13 Automated identification of PAOs in calibration dataset based on peak height. 149

Figure 6-14 Raman spectra numbers 55 and 78 from the calibration data set..... 150

Figure 6-15 Automated identification of PAOs in three test datasets (EDI100, NE100, SLO100) based on peak height. 151

Figure 6-16 Example misclassified spectra from data sets EDI100, NE100 and SLO100. 153

Figure 6-17 Automated identification of PAOs in calibration dataset based on PLS-DA. 155

Figure 6-18 Automated identification of PAOs in three test datasets (EDI100, NE100, SLO100) based on the developed PLS-DA model. 156

Figure 6-19 Example of a misclassified spectrum from the EDI100 dataset, and examples of spectra for samples lying close to the classification threshold in the calibration, NE100 and SLO100 datasets. 157

Figure 6-20 PLS-DA model statistics for all four Raman spectra datasets analysed for the presence or absence of polyphosphate. 158

Figure 7-1 Spectra of three pure compounds (sodium phosphate glass type 45 – ‘spg45’), PHB and glycogen) representing key EBPR metabolites and a typical bacterium spectrum, each normalised to the peak height of the most intense spectral band. Previously suggested bands to use as internal standards for normalisation of biological Raman spectra are highlighted in red. 172

Figure 7-2 Two examples of EBPR bacterial spectra overlaid with components 1 (representing polyphosphate character) and 4 (representing bacterial character) predicted using a 4-component MCR model built on the calibration data set used in Chapter 6. 173

Figure 7-3 SBR monitoring data acquired during the uptake-release test from which samples were taken for Raman spectroscopy. 178

Figure 7-4 Loadings for MCR model 1 (shown in red) and MCR model 2 (shown in blue). 179

Figure 7-5 Four candidate ‘direct’ (i.e. not normalised) measurements for determination of polyphosphate using Raman spectra. 181

Figure 7-6 Four candidate ‘reference’ measurements for normalisation of Raman spectral data for determination of polyphosphate. 182

Figure 7-7 Four candidate normalised measurements for the determination of polyphosphate using Raman spectra. 183

Figure 7-8 Residuals from MCR model 2 (MCR2) applied to Raman spectra of samples from SBR-D during the uptake release test described..... 185

Figure 7-9 Summed polyphosphate quantity measured by Raman spectroscopy of 50 individual cells for each sample taken during the uptake-release test..... 186

Figure 7-10 Raman spectra of aqueous polyphosphate solutions of varying concentration taken with two different acquisition times. 189

Figure 7-11 Schematic diagram of a bacterium sample on a CaF_2 substrate (left) and a liquid or aqueous sample enclosed in a sample well with a glass coverslip (right)..... 191

Figure 7-12 AFM image of EBPR bacteria dried on a glass microscope slide (top). Height data recorded from the three different line profiles indicated on the AFM image (bottom). 192

List of Tables

Table 3-1 Typical phases in a cycle run by a Sequencing Batch Reactor for Enhanced Biological Phosphorus Removal 24

Table 3-2 Chemical composition of synthetic wastewater used to feed the EBPR communities. 37

Table 4-1 Experimental parameters and FWHM determined from Raman line maps through 5 µm polystyrene beads in the lateral plane. 74

Table 4-2 Experimental parameters and FWHM determined from Raman line scans through different materials in the axial plane. 77

Table 5-1 Summary of background and baseline estimation methods investigated in this work 96

Table 5-2 Summary of selected PLS-DA model parameters and statistics 113

Table 5-3 Mean PLS regression model statistics for the Raman spectra of three concentration series of high molecular weight polyphosphate 117

Table 6-1 Position of major Raman spectrum peaks arising from a dried cell-pellet smear of Pseudomonas putida and an individual EBPR bacterium. 129

Table 6-2 Position of major Raman spectrum peaks arising from a pure compound (SPG45) and from polyphosphate within an EBPR bacterium. 131

Table 6-3 Position of major Raman spectrum peaks arising from pure PHB and from PHA within an EBPR bacterium. 133

Table 6-4 Position of major Raman spectrum peaks arising from pure glycogen and from glycogen within an EBPR bacterium. 136

Table 6-5 Variance in the position of the most intense Raman bands for each key metabolite across all bacterial samples analysed in this Chapter 6. 136

Table 6-6 Summary of the number of samples misclassified by two automated methods (peak height and PLS-DA) to identify candidate PAOs based on the presence of a polyphosphate signal in the Raman spectra. 159

Table 7-1 Overview of available methods for determination of intracellular polyphosphate 167

Table 7-2 LOD calculated for detection of aqueous polyphosphate with the Raman instrument, set-up and sample presentation described in this study. 190

Table 7-3 LOD of aqueous polyphosphate measured with Raman spectroscopy for two different acquisition times and expressed in different concentration units. .. 193

Chapter 1
Introduction

1.1 Phosphorus in Context 3
1.2 Overall Aim and Research Questions 6
1.3 Thesis Overview 7

1.1 PHOSPHORUS IN CONTEXT

Phosphorus is a vital, but limited, resource [1]. It is an essential nutrient required by all living organisms. It is necessary in fertilisers for the production of enough food to meet demand; and as an element, unlike energy, there is no substitute for phosphorus.

Phosphorus for use in fertiliser is mined from phosphate rock. Reserves of phosphate rock are highly geographically concentrated in politically unstable regions of the world (Figure 1-1). For example, 74% of mineable sedimentary reserves are located in Morocco and Western Sahara. Furthermore, contentious debate continues between stakeholders over just how much mineable phosphate rock remains globally. In the last seven years, estimates of how long the available phosphate rock reserves will last have ranged from as little as 50-100 years [2], to 300-400 years [3] in response to the substantial upwards revision of phosphate rock reserve figures for Morocco by the United States Geological Survey between 2009 and 2011 [4].

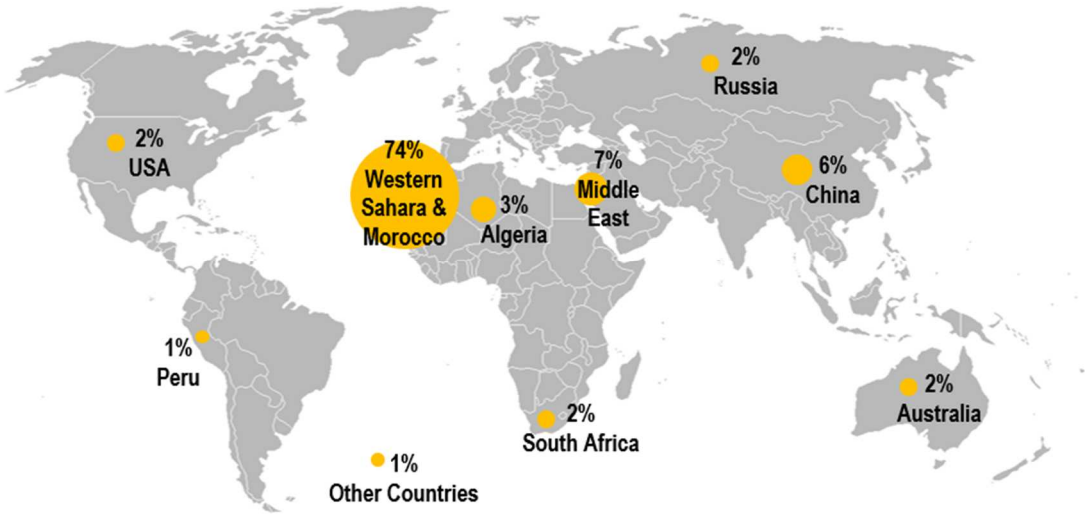


Figure 1-1 Geographical distribution of global phosphate rock reserves in 2015. Data taken and adapted from [5]

Whatever the true quantity of global phosphate rock reserves, the issue of availability of phosphorus on a national and regional level remains. In light of this and the challenge to provide food for a growing global population [6], the matter of 'phosphorus sustainability' has come to the fore [7], gaining increasing political attention at the regional, national and

international levels. Several platforms, including the European Sustainable Phosphorus Platform formed in March 2013, have emerged in a bid to address the challenge of phosphorus sustainability. In general these are designed to share knowledge and experience between stakeholders and to contribute to policy proposals. A useful overview of the current political landscape concerning phosphorus sustainability has been presented by Ulrich and Schnug [8]. This is a fast-developing field, however, and key developments have already occurred since this publication. For example, in 2014 the European Commission added phosphate rock to its list of 'critical raw materials' [9]. These materials are categorised as such when supply risk is high due to the fact worldwide production is concentrated in a few countries. The list is designed to incentivise new mining and recycling activities. With no sedimentary phosphate rock reserves in the EU, a concerted effort is needed to steward phosphorus effectively along the phosphorus 'chain' (Figure 1-2) and to 'close the phosphorus loop' by recovering phosphorus from wastewater at the end of the chain to reuse as fertiliser.

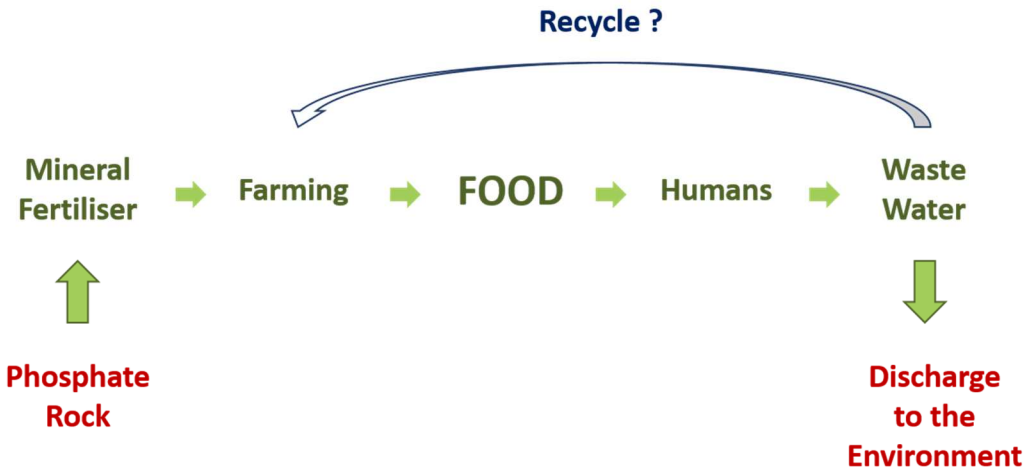


Figure 1-2 The phosphorus 'chain'. Phosphate rock is mined to produce fertiliser for the production of food. Humans consume the food and waste is passed out into wastewater. Wastewater is treated at sewage plants before discharge to the environment. Both ends of the chain can constitute an environmental challenge: supply of phosphate rock is at risk, and phosphorus in sensitive water bodies can be a pollutant causing eutrophication. The political drive to recycle phosphorus from wastewater back onto land as fertiliser is increasing.

Two basic methods exist for removal of phosphorus from wastewater: chemical and biological [10]. The chemical method uses aluminium or iron oxides to form phosphate precipitates which are then removed along with the wastewater treatment sludge. The biological method is known as Enhanced Biological Phosphorus Removal (EBPR) and this essentially engineers the microbial consortium present in the sewage sludge so that a certain type of bacteria is populous enough to be able to sequester phosphorus from wastewater [11]. Again, the phosphorus is ultimately removed from wastewater along with the sludge. These processes were developed in response to the fact that phosphorus, albeit a vital resource, is also a pollutant in water bodies sensitive to eutrophication [12]. The current employment of phosphorus removal technologies in wastewater treatment plants is generally driven by imposed regulatory discharge limits designed to manage the problem of eutrophication rather than phosphorus sustainability.

When considering not just the removal of phosphorus, but also its recovery from wastewater, EBPR constitutes a better platform than chemical precipitation [10]. The latter produces a relatively large amount of sludge containing high concentrations of metal ions. EBPR on the other hand, provides a better quality sludge for phosphorus recovery and reuse which contains much lower concentrations of metal ions. In addition it has the benefit of lowering operating costs which would otherwise be spent on dosing chemicals.

Technologies for the recovery of phosphorus from aqueous phase, sewage sludge and incineration sewage sludge ash have been summarised by Yoshida *et al.* [13]. In light of the changing political landscape surrounding phosphorus it could be conceived that recovery and not just removal of phosphorus from wastewater treatment plants could become incentivised. In addition to this, the adoption of phosphorus recovery technologies has been shown to solve other kinds of problem such as scaling [14]. For example, in 2010, Slough Wastewater Treatment Plant in the UK became the first site in Europe to partner with Ostara Nutrient Recovery Technologies (Canada) to install the Ostara Pearl® process [15, 16]. Downstream from EBPR, this process precipitates phosphorus from dewatering liquors as struvite, and this is sold as a slow-release fertiliser that Ostara brands 'Crystal Green®' [17]. For Thames Water this process modification solved a pre-existing problem of considerable scaling within plant pipework caused by struvite precipitating uncontrollably from the liquid waste stream [18, 19].

Whether employed solely to remove phosphorus from wastewater, or to form a platform from which phosphorus may be recovered and removed, optimising the performance and efficiency of the EBPR process has significant economic and environmental implications:

economic because if it fails, costs are incurred either in the form of back-up chemical dosing, or in the form of penalties imposed by regulatory bodies; and environmental because failure can cause eutrophication in receiving water bodies of the effluent. In the bigger political picture, the implications of EBPR performance in the context of recovering and reusing phosphorus should also be considered.

Despite its importance, operating EBPR to optimise performance and stability remains a challenge due to poor understanding of these bacterial ecosystems [20]. One particular difficulty in researching the process is that the functional EBPR organisms do not exist in pure culture. Research into the process has therefore had to be performed on mixed (and therefore highly variable) microbial cultures.

As a non-invasive, label-free, culture-independent technique capable of analysing live, single cells [21], Raman spectroscopy is a tool that can add extremely valuable insight into EBPR. Raman spectroscopy provides a biomolecular “fingerprint” of a sample based on characteristic vibrations of molecular bonds. In doing so, it has the potential to provide snapshots of bacterial type, physiology and function. Raman spectroscopy therefore has the potential to “watch” the dynamics of phosphorus accumulation and release in single cells as a function of time, and to relate this to cell physiology and external physicochemical (i.e. EBPR operating) conditions.

Raman spectroscopy has great potential for investigating EBPR but has been little used to date. Contributing reasons to this may include some of the following:

- Raman spectroscopy has a low profile in the context of EBPR research;
- Due to the scarcity of studies of applying Raman to EBPR research, there is a lack of standard methods for analysing the data;
- There is little understanding of what information is available from Raman spectra and how this could be used in EBPR research.

1.2 Overall Aim and Research Questions

In light of the limited application of Raman spectroscopy to EBPR research despite its advantages and potential to yield new insights to the process, the overarching aim of this thesis is to investigate the potential use of Raman spectroscopy to improve EBPR.

Specific research questions addressed in this work are as follows:

1. What is an effective, automated method for preprocessing whole Raman spectra of EBPR bacteria in preparation for exploratory analysis using multivariate statistics?
2. What information may be gained from qualitative analysis of Raman spectra of EBPR bacteria that could yield new insights to the process or open up new avenues of research?
3. How can the Raman spectra of EBPR bacteria be analysed to yield quantitative data?

In addressing these questions, it is hoped that the profile of Raman Spectroscopy as a tool for use amongst the EBPR research community is raised. Furthermore, through discussing spectral analysis methods in detail, it is intended that progress toward developing standard analysis protocols for this particular application are developed. Through emphasising the advantage Raman spectroscopy has for measuring cellular polyphosphate concentration and investigating ways to extract this data from Raman spectra, it is hoped that the quantitative capabilities of the technique are promoted. Finally, as an emerging technique for this particular application, it is intended that niche research areas where Raman Spectroscopy might be applied to EBPR research are highlighted.

1.3 THESIS OVERVIEW

The arrangement of this work is as follows:

Chapter two gives a brief overview of literature relating to EBPR and Raman spectroscopy applied to bacteria. It highlights the power and the potential of Raman spectroscopy to give new insights on EBPR which may lead to improved monitoring and control of the process.

Chapter three describes the design, build and validation of two sequencing batch reactors (SBRs) that were required for this work and which will be used in future studies by this research group. It also gives details of the methods used to monitor the performance of the SBRs.

Chapter four provides an introduction to the theory of Raman spectroscopy and describes the instrument and parameters used in this study. Furthermore, a brief description of the data analysis techniques applied to the Raman spectra is given.

Chapter five discusses the necessity and challenge of pre-processing Raman spectra to remove background and baseline effects. Several mathematical approaches for achieving this

are reviewed, discussed and applied to example spectra acquired from EBPR bacteria. Performance of the methods are compared with respect to visual appearance of the spectra and how they may affect subsequent qualitative and quantitative analysis. This is the first study of its type, investigating preprocessing of EBPR bacteria Raman spectra in detail. Furthermore it is one of very few studies comparing the application of a number of different baselining techniques to real data rather than synthetic data.

Chapter six presents a qualitative examination of 386 Raman spectra acquired from individual bacteria which were sampled from three different EBPR reactors. In line with a handful of previous studies, the capability of Raman spectroscopy to identify three metabolites key to the EBPR process – polyphosphate (PP), polyhydroxyalkanoates (PHA) and glycogen – in these bacteria is confirmed. Following this, the variability of the characteristic Raman signals for these compounds between samples was investigated in detail for the first time. The Raman spectra of EBPR bacteria are complex with many overlapping signals. Despite this, multivariate statistical tools have not yet been applied to these kind of samples. Here, for the first time, two multivariate analytical tools – MCR and PLS-DA – were applied to sample sets of Raman spectra of individual EBPR bacterial cells for the first time. The improvement of spectral analysis achievable with these tools was demonstrated and the potential for these tools to further EBPR research was discussed.

Chapter seven investigates the application of Raman spectroscopy to the quantification of polyphosphate in individual EBPR bacteria. In the few cases where this method has previously been used, the spectral analysis has typically related polyphosphate quantity directly to peak height. Here, the necessity of normalising spectra for quantitative analysis is discussed and for the first time, four different methods are investigated for normalising the Raman spectra of individual EBPR bacteria for the purpose of quantifying polyphosphate. Two of these methods are based on the use of MCR and to the author’s knowledge, this kind of analysis has not been applied previously to a normalisation problem such as this. Finally, an experimental estimation was made of the limit of detection of polyphosphate using the Raman instrumental set-up in this work.

Chapter eight summarises the findings of this thesis in light of the research questions posed at the beginning of this work.

Chapter 2
Literature Review

2.1 ENHANCED BIOLOGICAL PHOSPHORUS REMOVAL 11
2.2 RAMAN SPECTRA OF BACTERIA 15
2.3 RAMAN SPECTRA APPLIED TO EBPR 18

2.1 ENHANCED BIOLOGICAL PHOSPHORUS REMOVAL

Enhanced Biological Phosphorus Removal (EBPR) is a biological treatment process for removing excess phosphorus from wastewater [22]. Essentially, it is a modified activated sludge process that was developed after the first, somewhat providential, observations of phosphorus removal in conventional activated sludge plants in the 1970s [23].

EBPR works because of putative microorganisms known collectively as polyphosphate accumulating organisms (PAOs). Every living organism needs phosphorus to survive, but PAOs are bacteria that are distinctive in their ability to assimilate and store more phosphorus than they need for immediate growth. The excess phosphorus is stored as intracellular polyphosphate – a linear polymer comprising tens to hundreds of phosphate residues [24]. During the EBPR process, phosphate is thus concentrated into the biomass and subsequently removed from wastewater by continuously 'wasting' (removing) a fraction of the activated sludge.

PAOs are not one particular species of bacteria, but it is understood from research to date that they comprise several different species, perhaps not all of which have been identified [25]. To date, none of the recognised 'candidate' PAO species has been successfully isolated, grown and maintained in pure culture. The successful function of EBPR thus relies on managing process conditions to favour the growth of PAOs at the expense of other microorganisms. In this way the PAO population is enriched in the total microbial community and their collective function becomes sufficient to remove all the available phosphate from the influent wastewater.

EBPR is a mature technology, operated globally with a variety of configurations. Understanding of the scientific principles underlying EBPR has improved considerably during the past 40 years and is the subject of several comprehensive reviews [11, 20, 22, 25, 26]. In many cases, however, the process remains operationally unstable, and can fail unexpectedly [27]. Application of tools such as Raman spectroscopy to EBPR could yield new insights and ideas for improving process performance.

Giving PAOs a Selective Advantage

It is widely accepted that the key condition required to enrich PAOs in EBPR is to cycle between anaerobic and aerobic conditions. The 'ordinary heterotrophic organisms' (OHOs)

in activated sludge require a simultaneous supply of organic carbon and oxygen in order to survive and grow. PAOs tend to be slow-growing with respect to OHOs but their ability to make intracellular stores of both carbon compounds (in the form of polyhydroxyalkanoates – PHAs) and polyphosphate mean they have a niche metabolism that may be exploited to give them a competitive edge for survival. In an EBPR process, the organic carbon is introduced to the biomass under anaerobic conditions. PAOs essentially assimilate and hoard the available short-chain carbon compounds (i.e. volatile fatty acids – VFAs) during this phase of operation making them unavailable to OHOs during the subsequent aerobic phase.

The principal biochemical transformations occurring during EBPR are summarised in Figure 2-1. Under anaerobic conditions, PAOs break down polyphosphate and glycogen stores to produce energy and reducing power. Phosphate is released into solution as a byproduct of this transformation. The energy and reducing power produced are used to assimilate VFAs (including acetate) and to store them as polyhydroxyalkanoates (PHAs). During the subsequent aerobic conditions, PHAs are respired for cell growth and for replenishment of polyphosphate and glycogen stores. The mass of phosphate taken up from solution to replenish polyphosphate stores is greater than that released during the anaerobic stage so that there is a net uptake of phosphorus during an EBPR cycle.

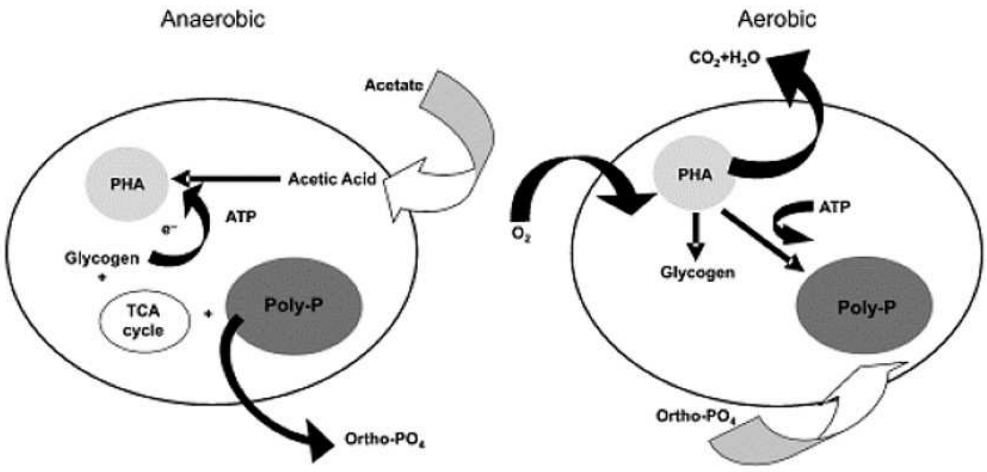


Figure 2-1 Principal biochemical transformations occurring in PAOs during EBPR under anaerobic (left) and aerobic (right) phases. Reproduced from [28].

The expected trends in concentrations of key compounds during anaerobic and aerobic phases of the EBPR are summarised in Figure 2-2. During an anaerobic phase, concentrations of VFAs (including acetate) decrease rapidly to very low levels. At the same time, concentrations of phosphate (PO_4-P) in solution increase as PAOs break down polyphosphate and release the phosphate residues. At the beginning of the aerobic phase, any residual COD is rapidly consumed. Throughout this phase the concentration of phosphate in solution decreases as phosphate is re-assimilated by PAOs. In a successful EBPR reactor, the concentration of phosphate at the end of the aerobic phase (i.e. in the treated effluent) is less than the regulatory discharge limit imposed on a treatment plant (or alternatively a target concentration in a research setting).

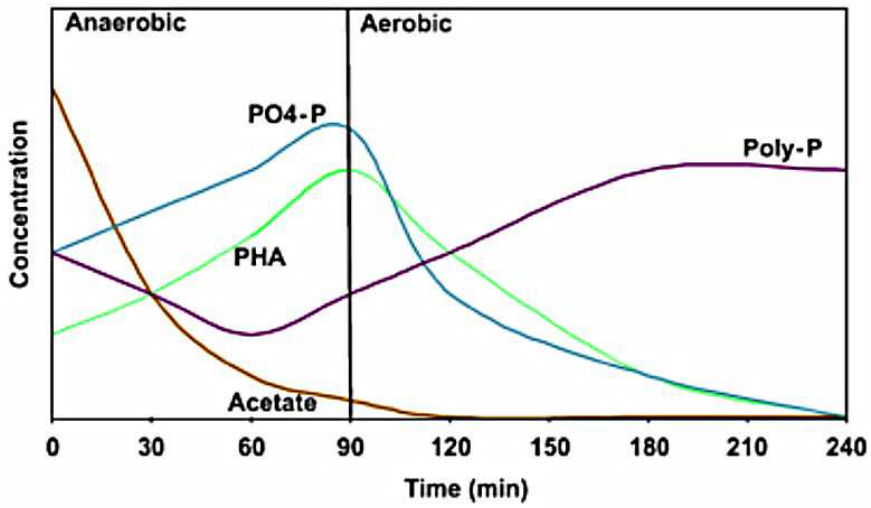


Figure 2-2 Concentration profiles of key compounds during the anaerobic and aerobic phases of EBPR. Reproduced from [28].

The biochemical transformations occurring in PAOs during the EBPR process are not monitored directly in general practice. Rather, the concentrations of phosphate and biological oxygen demand (BOD) or chemical oxygen demand (COD) in solution are measured as indicators of process performance. BOD and COD are tests that determine how much oxygen is consumed by a sample to oxidise the carbon compounds present. BOD and COD tests thus indirectly measure the concentration of carbon compounds.

Besides PAOs and OHOs, the third (and final) major group of organisms included in EBPR models [29] is glycogen accumulating organisms (GAOs). GAOs behave similarly to PAOs

in the respect that they can also assimilate VFAs under anaerobic conditions and store them as intracellular PHAs. Whereas PAOs break down polyphosphate and glycogen stores to provide the energy to assimilate VFAs, however, GAOs use just glycogen stores to achieve this since they do not have the ability to store polyphosphate. GAOs thus have the ability to compete with PAOs for carbon compounds during the anaerobic phase of EBPR but do not have the ability to remove phosphate from wastewater beyond their immediate growth requirement. Proliferation of GAOs in EBPR populations have been blamed for process deterioration in some cases and thus where possible, operating conditions are managed to give PAOs a competitive edge over GAOs [30]. For example, ambient to cool temperatures and a pH between 7.25 and 8, are thought to favour PAO growth and function over that of GAOs [31, 32].

The identity of some PAOs is known. Hesselmann *et al.* used Fluorescence in situ Hybridisation (FISH) combined with DAPI staining to identify *Rhodocyclus*-related *Betaproteobacteria* as a PAO candidate in two lab-based sequencing batch reactors (SBRs) for EBPR [33]. It was given the name *Candidatus* Accumulibacter phosphatis and has subsequently been detected in many laboratory and full-scale systems [34, 35]. It seems likely that this species constitutes a major PAO population in EBPR processes [36], however, other species such as *Tetrasphaera*-related *Actinobacteria* have been also been recognised as exhibiting the phenotypic behavior of a PAO [37, 38]. More recent studies on PAOs have reported that there are several clades within the recognised PAO species [39] and that some PAOs, termed denitrifying polyphosphate accumulating organisms (DPAOs), can use nitrate instead of oxygen as an electron acceptor [40]. Furthermore, some have suggested that PAOs and GAOs could be the same organisms that switch between metabolising polyphosphate or glycogen depending on operating conditions [41]. Understanding the behavior of PAOs is difficult in mixed culture. Analysis techniques typically measure ‘bulk’ parameters that represent the overall behavior of a bacterial community. The fact that Raman spectroscopy can analyse single, live cells, means it would perhaps have the potential to be able to measure behaviours like metabolic switching in vivo.

The phosphate removed by EBPR may subsequently in some cases be recovered by various methods including direct application to land as biosolids, biological release and precipitation of phosphorus, and incineration followed by phosphorus recovery [42]. In the context of the phosphorus sustainability issue discussed in Chapter 1, another stream of research related to EBPR is investigation of ways to release and recover polyphosphate from PAOs [43]. This is another area where Raman spectroscopy could be applied.

Since EBPR processes can fail unexpectedly, and since PAOs do not yet exist in pure culture, there remains much to be learned about EBPR in order to optimise its performance and stability. Raman spectroscopy has great potential to aid exploratory analysis of this microbial system. Examples of previous Raman spectroscopic studies on bacteria are presented below.

2.2 RAMAN SPECTRA OF BACTERIA

Raman spectroscopy is a rapid, non-invasive technique that can be applied to a wide variety of sample-types. It provides a biochemical 'fingerprint' of the subject based on characteristic vibrations of chemical bonds. A more detailed description of the technique is given in Chapter 4. By coupling a Raman spectrometer to a confocal microscope, this method may be used to interrogate microscopic-scale samples such as bacteria. Some use the term 'Raman microscopy' [44] or 'Raman microspectroscopy' [45] to describe this, whereas others continue to use the term 'Raman spectroscopy' [46]. Due to the variety of terms used in the literature, it should be noted that in this work the term Raman spectroscopy will be used to describe the analysis of bacteria via a Raman spectrometer coupled to a confocal microscope.

Raman scattering is an inherently weak effect and a number of advanced Raman spectroscopy techniques have been developed to overcome this limitation. These include Coherent Anti-Stokes Raman spectroscopy (CARS), Stimulated Raman Spectroscopy (SRS) and Surface-Enhanced Raman Spectroscopy (SERS) [47]. These advanced Raman techniques are not investigated in this work for the following reasons. CARS and SRS are both multi-photon techniques that employ two incident lasers. The difference in wavelength between these two lasers is matched to a particular vibrational frequency of the sample. This causes stimulated excitation of the selected molecular vibration and thus increases the intensity of Raman scattering. In CARS, new radiation at the anti-stokes frequency is detected [48] whereas in SRS, changes in intensity of the excitation beams are detected [49]. Since CARS and SRS effectively target one particular Raman signal rather than a spectral range, they are typically used in imaging applications i.e. to map changes in that signal over a defined sample area [50]. In this exploratory work, full Raman spectra of the EBPR bacteria are desired rather than imaging of one or two selected signals. SERS uses noble metal surfaces (e.g. gold or silver) with nanoscale features (e.g. nanoparticles, nanoscale-

patterned surfaces) to enhance electromagnetic energy in the vicinity of the target sample. This in turn enhances the intensity of Raman-scattered light arising from that sample [51]. One difficulty in using SERS to characterise bacteria, however, is that SERS signal intensity is a function of distance from the surface of the metal. Bacterial size tends to be in the order of microns but signal enhancement using SERS substrates has been found to decay considerably within 10 nm from the surface of the metal. Another possible problem with SERS is that sometimes the signal is not reproducible so that different spectra can be produced with each measurement [47]. For these reasons, conventional (spontaneous) Raman spectroscopy is the technique used in this particular work.

The strengths of Raman spectroscopy are substantial in relation to the study of biological matter. It is rapid, non-invasive, requires little or no sample preparation, contains a wealth of information, is non-targeted, can interrogate small samples down to the single-cell level, and can be applied to live cells. Some examples of studies benefitting from these advantages in the field of microbiology are highlighted below.

The application of Raman spectroscopy to microbial identification has been investigated in a number of studies. In principle this uses the idea of the Raman spectrum as a fingerprint that contains enough biochemical information to make it unique to each microbial 'type'. It can thus potentially be used to identify microbial species [52]. This is of particular interest in a clinical setting where the rapid diagnoses of pathogenic bacteria either from a patient or from the hospital environment can be critical to prompt treatment or the prevention of hospital-acquired infections [53]. Conventional tests for the identification of pathogenic microorganisms can take up to 2-3 days whereas a Raman-based identification could potentially be completed within 6 to 8 hours after starting a culture [54]. Microbial typing by Raman spectroscopy down to the strain level has been demonstrated in *Staphylococcus aureus* [52], *Staphylococcus epidermis* [55], *Mycobacterium tuberculosis* [56], *Escherichia coli* / *Legionella pneumophila* [57], and *Bacillus subtilis* [58]. These were all essentially proof-of-principle studies and for commercial development would require extension and confirmation with larger test sets. They do, however, demonstrate the potential power of Raman spectroscopy to detect very small microbial differences on a very short timescale.

Ciobotă *et al.* [59] produced the first study testing Raman-based typing in bacteria known to produce the storage polymer polyhydroxybutyrate (PHB). Investigating the typing of *Bacillus megaterium* DSM 90, *Bacillus thuringiensis* (DSM 530, DSM 5725, ATCC10782), *Azohydromonas lata* DSM 1122, *Cupriavidus necator* DSM 428 and *Acidiphilum cryptum* JF-5, Ciobotă *et al.* found that in different growth phases, the bacteria contained different

amounts and forms of PHB. In the lag phase the Raman spectra showed no signals for PHB. In the exponential phase, the PHB present seemed to be amorphous, and in the stationary phase, the PHB had a spectral signature consistent with crystalline PHB. In lag or exponential phase the identification of these species by Raman spectroscopy was successful. In stationary phase, however, the crystalline PHB Raman signal dominated the spectra to the extent that the different bacteria species could not be distinguished. In EBPR, three storage compounds – polyphosphate, PHAs (including PHB) and glycogen – are accumulated by the key bacteria in the process. Based on the results of this study, it could be implied that the potential for typing different kinds of PAOs or GAOs with Raman spectroscopy is limited.

A number of studies have been applied to studying the composition and relative amounts of intracellular components in microbial species. De Gelder *et al.* [60] investigated relative changes in the Raman spectra of *Cupriavidus metallidurans* LMG 1195 over a life cycle with the aim of relating spectral changes to information on changes in cell components. Similarly, Beer *et al.* [61] studied metabolite composition in single cells of the unculturable snow alga *Chlamydomonas nivalis* over a life cycle with the purpose of investigating secondary metabolites that are likely related to adaptation of the organism to stress. The power of Raman spectroscopy to interrogate single cells was particularly useful in this study because this particular alga is unculturable. Watanabe *et al.* [62] studied iron compounds in cell fractions of magnetotactic bacteria, using Raman as a tool to investigate the intermediary compounds of magnetosome formation. This study did not employ, however, the full capabilities of Raman spectroscopy to potentially study this process in situ.

One difficulty in analysing live, single, non-adherent cells is to hold them still (or at least within the laser spot) whilst taking a measurement. This has been overcome in some cases by using a Raman system with an optical trap. Avetisyan *et al.* [63] used this method to investigate the uptake of trehalose by live, single cells of *Sinorhizobium meliloti*, demonstrating that carbohydrate uptake and metabolism could be monitored in real time. These kind of online studies could be employed for the in-situ direct monitoring of components produced by biological processes such as biofuels and bioplastics. Wu *et al.* [64] used Raman spectra to evaluate lipid profiles (including composition, chain length, degree of unsaturation and chain length) in optically-trapped single cells of several algal species that are candidates for biofuel production. De Gelder *et al.* [65] reported that Raman spectroscopy was a suitable technique for monitoring the relative quantity of PHB content in *Cupriavidus metallidurans* LMG 1195. Although the latter experiment was carried out on dried bacterial samples, it demonstrated the potential for monitoring intracellular production

of bioplastics. Furthermore, PHB is one of the key metabolites in EBPR and monitoring this in relation to process conditions could enhance both monitoring and control of phosphate removal performance and stability.

Surprisingly few studies have used the power of Raman spectroscopy as a technique to investigate unculturable bacteria in natural communities. Vitek *et al.* [66] studied endolithic communities in gypsum crust from the Atacama Desert. They demonstrated that Raman spectroscopy could be used to distinguish between prokaryotic cyanobacteria and eukaryotic algae in situ based on the pigment composition of these organisms. Huang *et al.* combined Raman spectroscopy with Fluorescence in situ Hybridisation (FISH) to observe the incorporation of ^{13}C into single microbial cells. The feasibility of this technique for conducting structure-function analyses of complex microbial communities was demonstrated [67].

Several examples of the application of Raman spectroscopy to study bacteria have been summarised in this review. From these it can be seen that this technique has great potential for exploratory investigation of the EBPR process. Despite this, Raman spectroscopy has been applied to this field in only a handful of studies. These cases are discussed below.

2.3 RAMAN SPECTRA APPLIED TO EBPR

Majed *et al.* reported the first investigation involving the application of Raman spectroscopy to the EBPR process, or even to analysing intracellular polyphosphate [68]. The main purpose of this study was to develop a Raman-based method to both identify and quantify polyphosphate in individual EBPR cells. Analysing individual EBPR cells dried on a calcium fluoride (CaF_2) substrate, it was observed that characteristic polyphosphate Raman bands were clearly recognisable in the bacterial spectra, albeit shifted in position with respect to the spectra of pure, short-chain (≤ 6 phosphate residue) polyphosphates. Based on the premise that the polyphosphate peak heights were proportional to polyphosphate concentration, the quantity of polyphosphate in individual cells was investigated as a function of time through one anaerobic-aerobic EBPR cycle. It was shown that the intracellular levels of polyphosphate within individual cells at each time point varied i.e. was heterogeneous, but the summed intensity per sample was inversely correlated with the bulk orthophosphate concentration measured in solution suggesting population level information

was also valid. The potential for using Raman spectroscopy to observe intracellular polyphosphate dynamics during the EBPR process was thus demonstrated.

Since this pioneering work, a few more studies have used Raman spectroscopy to investigate polyphosphate in biological samples. For example, Omelon *et al.* studied the role of polyphosphate in apatite biomineralisation in the skeletons of elasmobranch fishes [69], Bednárová *et al.* investigated the presence and quantity of polyphosphate in the vacuoles of living yeast cells of *Candida albicans* [70], and Eder *et al.* [71] used Raman spectroscopy to investigate magnetite crystals and sulphur compounds in magnetotactic bacteria but in doing so also identified polyphosphate in the bacteria by its characteristic spectral bands.

The second study applying Raman spectroscopy to EBPR bacteria extended the analysis to include PHA and glycogen as well as polyphosphate [72]. This was not the first study to measure PHAs [59, 65] or glycogen [73] in bacteria with Raman spectroscopy. It was, however, the first report showing that all three metabolites could be identified and quantified simultaneously in EBPR bacteria. This work was later extended to suggest candidate PAOs and GAOs could be identified based on the particular combination of metabolites present in their Raman spectra [74]. The potential for quantifying these metabolites in distinct PAO and GAO populations rather than across both populations was thus demonstrated. This could be useful particularly for glycogen which is accumulated in both PAOs and GAOs but conventional methods for extracting and measuring glycogen would give results representing the combined behaviour of two different populations.

In these few studies the potential application of Raman spectroscopy to EBPR was demonstrated but the spectral analysis was based simply on peak height. Theoretically, peak height in a Raman spectrum is proportional to the number of scattering molecules giving rise to that peak. It is also practically affected, however, by factors such as laser wavelength and power, optical alignment, sample presentation and fluorescence (discussed further in Chapters 4 and 7). Further studies building on the early work of Majed *et al.* are thus warranted to investigate ways of standardising Raman-based characterisation of EBPR bacteria as well as explore how this characterisation may be used to improve the EBPR process.

Chapter 3
Methods I: Design, Build and Validation of Sequencing Batch Reactors for Enhanced Biological Phosphorus Removal

3.1 REQUIREMENT FOR SEQUENCING BATCH REACTORS 23
3.2 DESIGN AND BUILD OF SEQUENCING BATCH REACTORS 25
3.2.1 Specifications 25
3.2.2 Design and Build 26
3.3 METHODS FOR MONITORING 32
3.3.1 Determination of Orthophosphate 32
3.3.2 Determination of Chemical Oxygen Demand (COD) 33
3.3.3 Determination of Suspended Solids 33
3.3.4 Measurement of Dissolved Oxygen (DO), pH and Temperature 34
3.3.5 Neisser Staining for Microscopy 34
3.3.6 Scanning Electron Microscopy (SEM) 34
3.4 VALIDATION OF SEQUENCING BATCH REACTORS 35
3.4.1 Operating Conditions 35
3.4.2 Performance 44
3.4.3 Second Run of the SBRs 46
3.4.4 Summary of Modifications made during SBR Operation 51
3.5 FUTURE IMPROVEMENTS FOR SEQUENCING BATCH REACTORS 52

3.1 Requirement for Sequencing Batch Reactors

To date, Polyphosphate Accumulating Organisms (PAOs) are still considered to be unculturable bacteria. As such they are not available to purchase from microbial culture collections such as the American Type Culture Collection (ATCC). Furthermore, even if they were available, the know-how to sustain these organisms in pure culture is not yet available despite many decades of research. For experimentation, PAOs must thus be sampled from mixed microbial communities – either from the natural environment or from systems engineered to enrich (selectively increase the abundance of) PAOs within the community.

For this work, practical options for obtaining PAOs included requesting samples from full-scale Enhanced Biological Phosphorus Removal plants or from research groups already working with these bacteria. Alternatively, an experimental rig (bioreactor) could be built to develop and maintain an enriched community of PAOs in-house. It was perceived that working with samples provided by other groups would have a number of problems: Firstly, the bacteria would inevitably undergo perturbation during transport, particularly with respect to growth rate and metabolic state. Secondly, research groups have their own aims and objectives, so little freedom could be expected in defining conditions under which samples were to be taken. Thirdly, samples provided from full-scale EBPR plants would have been subject to highly variable operating conditions and furthermore, information on those conditions may not be available. Therefore, for the sake of having PAOs readily available from well-defined conditions specified by us, it was decided that an experimental rig for EBPR should be built in-house.

As discussed in more detail in Chapter 2, EBPR relies on cycling the mixed microbial community through anaerobic (feast) and aerobic (famine) phases. PAOs have the ability to store levels of polyphosphate and carbon compounds beyond their immediate requirement for growth. Fresh influent is introduced at the beginning of the anaerobic phase when ordinary heterotrophic organisms are unable to assimilate volatile fatty acids (VFAs). By contrast, PAOs are able to use the energy from stored polyphosphate to assimilate the VFAs and later metabolise them during the aerobic phase. This gives them a selective advantage under this feast-famine regime and leads to their enrichment within the microbial community.

The experimental rig for EBPR needed to provide the cycling anaerobic-aerobic phases is described above. In practice these phases could be separated either in space by a continuous

flow reactor, or in time by a sequencing batch reactor (SBR). Typical phases in an SBR cycle for EBPR are shown in Table 3-1. Both reactor configurations can perform EBPR equally well when nitrification is inhibited (as it was in these experiments). Automated control of an SBR is more complex than that of a continuous flow reactor. Operation of the former requires a timed sequence to be controlled automatically so that mixing, aeration and pumping all occur at the correct point in the process. In continuous flow reactors, mixing, aeration and pumping are all continuously switched on. Control of the length of anaerobic and aerobic phases in a continuous reactor, however, is via the original design of the reactor (capacity, dimensions, etc.) and by altering the wastewater flow rate. In SBRs, process parameters such as the length of the anaerobic and aerobic phases are much easier to change by programming the timed sequence accordingly. Furthermore, no wastewater shortcuts the process in an SBR whereas variability in residence time of wastewater can be expected in a continuous flow reactor due to the existence of different flow patterns. For adaptability and simplicity of design, therefore, we opted to build SBRs.

Table 3-1 Typical phases in a cycle run by a Sequencing Batch Reactor for Enhanced Biological Phosphorus Removal

SBR PHASE	DESCRIPTION
Fill	A specified volume of influent is pumped into the reactor tank
Anaerobic	The contents of the reactor tank are mixed under anaerobic conditions
Aerobic	The contents of the reactor tank are mixed under aerobic conditions
Waste	A small volume of mixed liquor is drawn off in order to control biomass levels in the reactor
Settle	Mixing is stopped and the solids in the reactor are allowed to settle
Draw	A specified volume of supernatant i.e. effluent is drawn from the reactor
Idle	All pumps, mixers and aerators are off i.e. idle, creating a pause before the beginning of the next cycle

It was decided that two SBRs would be built and run in parallel. The two microbial communities in them would be allowed to develop and adapt independently of each other (i.e. each reactor had its own pipework and care was taken not to cross-contaminate bacteria from one reactor to the other). There were several reasons for building two SBRs in this way:

(i) if one reactor should fail (i.e. the EBPR performance deteriorate and the PAO population diminish), the other reactor would act as a back-up; (ii) if operated under similar conditions, PAOs from the two reactors could be compared with each other; (iii) if operated under different conditions, PAOs from the two reactors could be contrasted with each other. The design, build and validation of these SBRs is the subject of the remainder of this chapter.

3.2 DESIGN AND BUILD OF SEQUENCING BATCH REACTORS

3.2.1 Specifications

Purpose

The primary purpose of the SBRs was to maintain and allow experimentation with two independent, enriched cultures of PAOs in the laboratory for up to one year. The SBRs were required, therefore, to cycle reliably and continuously through the fill-anaerobic-aerobic-waste-settle-draw-idle sequence for up to twelve months.

Automation

Automated operation of the SBRs was necessary since the reactor sequence needed to run continuously. For effective operation and the freedom to experiment, the ability to alter the programmed length of each phase in the sequence was required. In addition, from a practical perspective, it was desired that the SBRs were able, if necessary, to run independently for up to one week without any intervention.

Intervention

For the sake of maintenance and for troubleshooting any operational problems, the ability to override the automated sequence (i.e. to be able to turn individual components on and off outwith the automated program) was required.

Monitoring

Continuous monitoring and logging of key parameters – pH, dissolved oxygen (DO) and temperature – were required to ensure that the SBRs were working within the required conditions and to highlight any operational problems. In addition to this, a sampling port was required in each reactor to permit easy sampling for further monitoring and testing.

Maintenance

Microbial growth and accumulation in pipes and on sensors was a concern with respect to maintenance of the reactors. On sensors, biomass accumulation could affect the accuracy of readings taken. In pipework, particularly on the influent side, biofilms could consume carbon sources and nutrients, thereby affecting the composition of synthetic wastewater before it entered the reactors. In light of this, it was necessary that it was easy to access the sensors at least once per week to clean them, and that it was easy to either clean or replace the influent pipework periodically.

3.2.2 Design and Build

Overview

A schematic diagram of the overall SBR layout is shown in Figure 3-1. For the two glass reactor tanks (containing the enriched cultures of PAOs), a working volume of 2 L was selected for practical purposes: small enough for a single operator to handle, but large enough to allow sufficient mixed liquor volumes to be removed periodically for experimental and monitoring purposes.

Each reactor tank was equipped with a mechanical mixer, two air lines and a float sensor to prevent overflow. For continuous monitoring, each reactor tank also contained a thermocouple, pH sensor and DO sensor. All these components, together with the influent and effluent lines for each tank were mounted on a 10 mm thick transparent (acrylic) top-plate. This plate was supported on an aluminium structural frame (Bosch Rexroth Extrusion), which itself rested on the laboratory benchtop. Photographs depicting the SBR layout are shown in Figure 3-2.

The two reactor tanks were seated on a scissor stand underneath the acrylic top-plate. When the scissor stand was extended, the rig was in operational configuration and the top edge of each reactor tank nested within a circular recess in the underside of the acrylic top-plate. When the scissor stand was contracted, the two reactor tanks could be removed easily from the rig so that pipework and sensors could be cleaned and maintained.

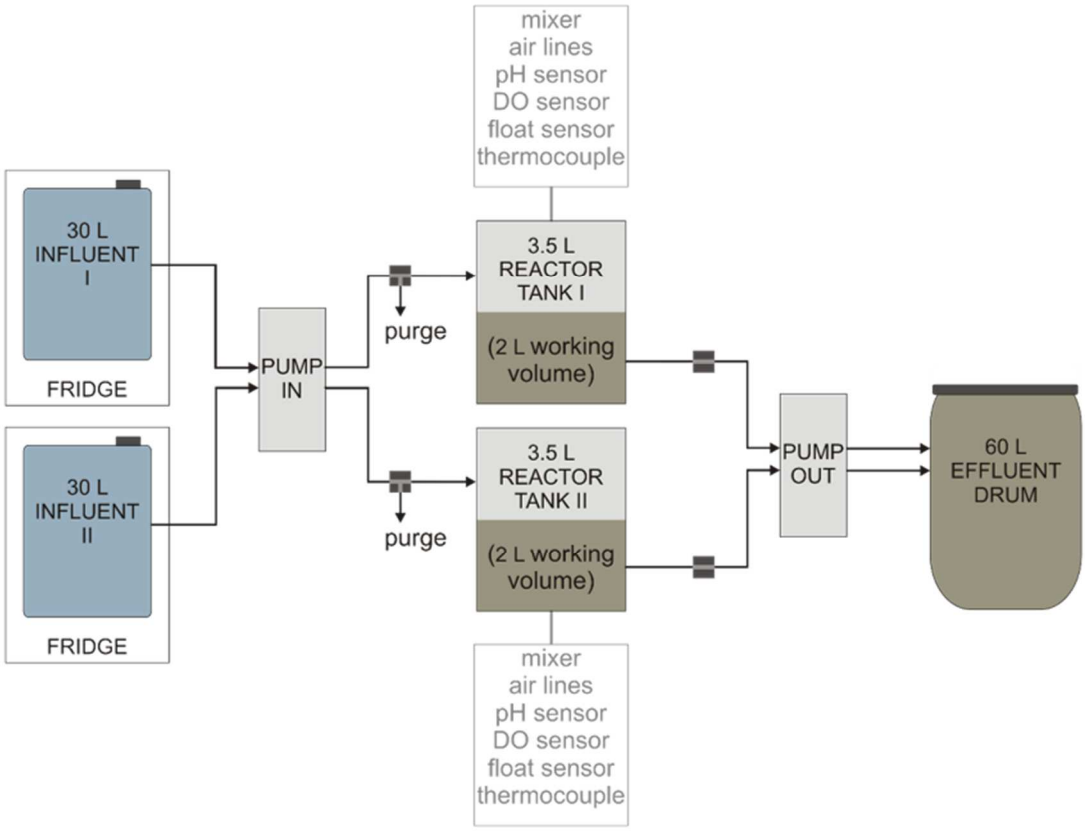


Figure 3-1 Schematic diagram depicting the layout of the Sequencing Batch Reactors

During the fill phase of each cycle, influent (synthetic wastewater) was pumped into each SBR from two separate 30 L reservoirs. These reservoirs contained enough influent for continuous operation of the SBRs for one week. During the waste and draw phases of each cycle, mixed liquor and supernatant (i.e. effluent) respectively were pumped from each SBR into a single 60 L drum. This drum had the capacity to hold waste (effluent) from continuous operation of the SBRs for one week.

The design and components of the SBRs are discussed in more detail below. A full list of components, together with suppliers and parts numbers where available, are provided in Appendix 1.

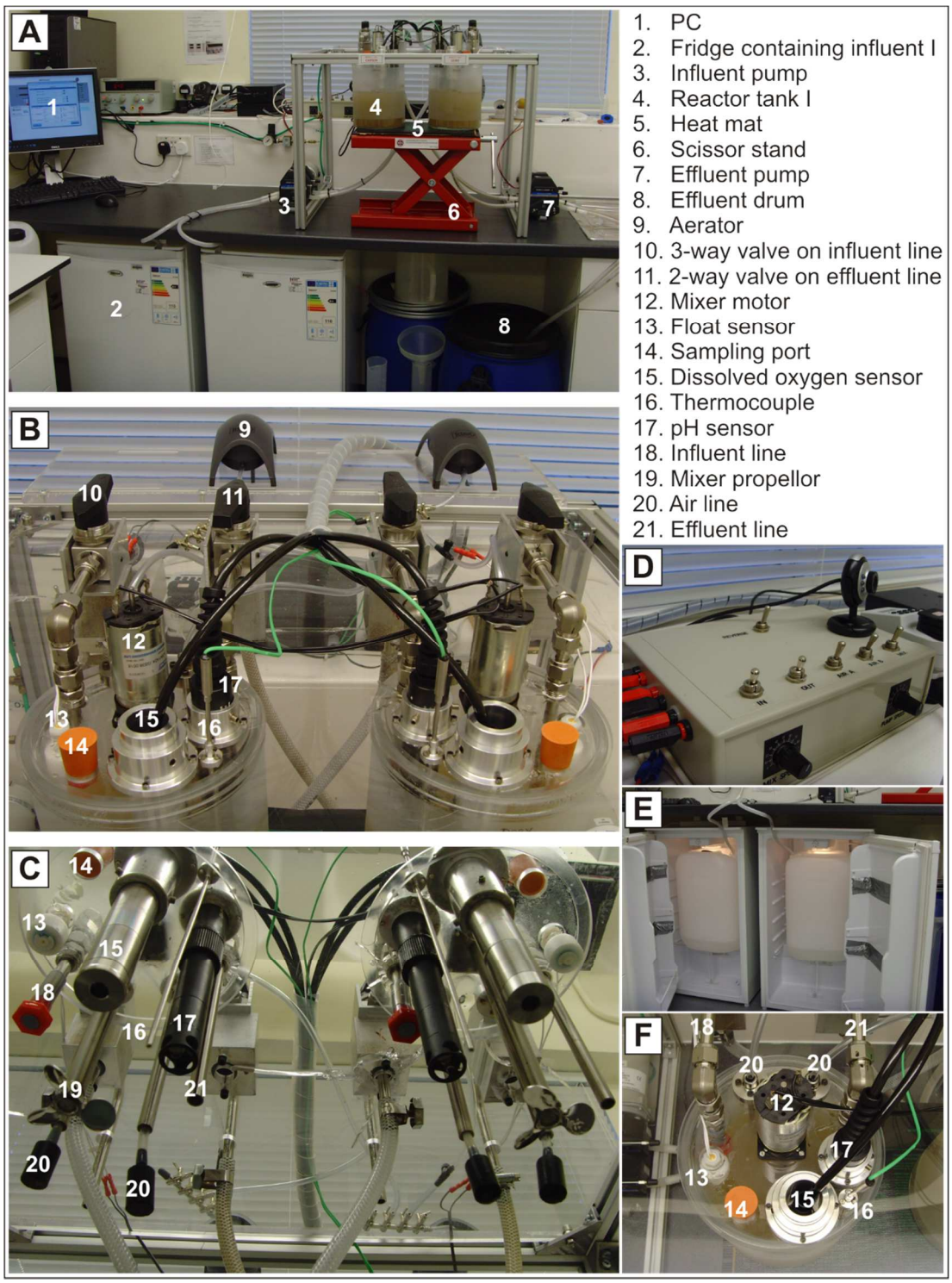


Figure 3-2 Photographs depicting the SBR layout (A), top of the reactor tanks (B), view from underneath the acrylic top-plate with the tanks removed (C), control box and webcam (D), open fridges containing 30 L influent containers (E) and plan view of reactor I (F).

A schematic diagram of the control and monitoring interface of the SBRs is shown in Figure 3-3. Automated operation of the SBR cycles was facilitated by a LabVIEW (National Instruments) program written in-house. This program was designed to operate certain components (reactor pumps, mixers and aerators) at the required times during each SBR cycle (Figure 3-4). The interface between the LabVIEW program and the components was provided by a data acquisition (DAQ) device and a control box that was designed and constructed in-house. The control box comprised relay switches (interfaces for programmed on/off control) as well as mechanical switches to override the program and allow the components to be switched on manually. The control box also provided manual speed control for the mixers and pumps. Overriding the automated control to switch components off manually required either that the operation program was stopped or that the power to the 24 VDC power supply was switched off. The heat mat was a later modification made to the SBRs and is explained in section 3.4.4.

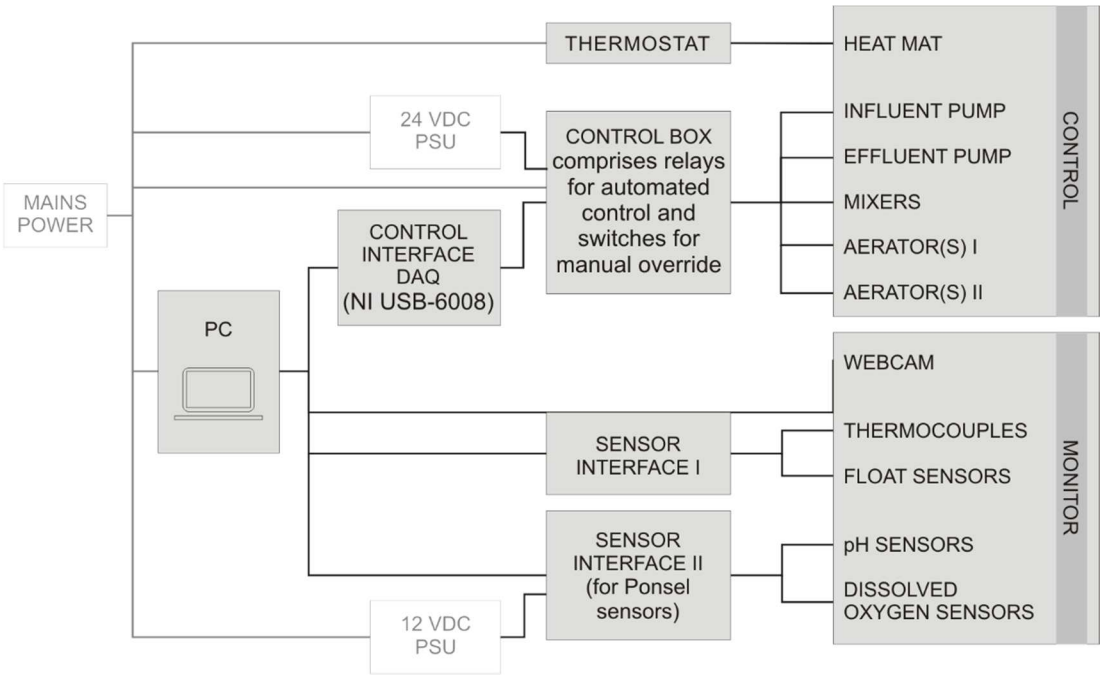


Figure 3-3 Schematic diagram of the control and monitoring interface of the Sequencing Batch Reactors. 12- and 24- VDC PSU indicate 12 and 24 Volt direct current power supply units respectively. DAQ stands for data acquisition device.

HARDWARE	PHASE						
	fill	react - anaerobic	react - aerobic	sludge wasting	settle	draw	idle
Influent Pump (2 heads)							
Effluent Pump (2 heads)							
Aerators I and II							
Mixer motors I and II							
Liquid Level sensors I and II							

Instrument OFF

Instrument ON

Figure 3-4 Components operational during each phase of the Sequencing Batch Reactor cycle

SBR Monitoring

For monitoring pH and dissolved oxygen (DO), sensors manufactured by Ponsel were selected on the basis that they are robust, have waterproof housing (IP68) and are designed especially for continuous environmental monitoring. Optical, rather than galvanic DO sensors were selected on the basis that they would provide an accurate reading even during the settle phase of SBR operation. By contrast, galvanic sensors require stirring to obtain an accurate measurement. It was intended that the sensors would be interfaced with the developed LabView control program so that feedback-based control of dissolved oxygen levels would be possible. At this stage, however, this was not achieved, but remains a target for future development. The Ponsel sensors required an external 12 V power source and were interfaced with the PC using Modbus RS-485 protocol. Data was logged with the manufacturer's software Calsens. For monitoring that required samples to be taken from the reactors, the SBR design included a sampling port on the top of each reactor for ease of access.

Pumping, Pipework and Reservoirs

Two peristaltic pumps were employed with enough torque to drive two pump heads each. One pump supplied each of the two reactor tanks with influent, and the other pump drew effluent from the reactor tanks. 24V DC pumps were selected on the basis that they are economical and simple to control with respect to AC pumps. A combined control for pumping speed of both influent and effluent pumps was installed on the reactor control box. Pipework with 3/8" diameter was selected to minimise the likelihood of blockage from

bacterial flocs. 3-way valves were installed on the influent side of the reactor tanks to allow influent lines to be cleaned and purged. 2-way valves were installed on the effluent side of the reactor tanks to prevent potential siphoning of the reactor contents whenever the pump-clamp on the effluent tubing was released. The influent and effluent reservoirs were designed to have enough capacity that the SBRs could run for a week without attendance. The influent reservoirs were stored in fridges to retard bacterial growth in the synthetic wastewater. The fridges were modified to be able to contain the 30 L influent reservoirs.

Mixing

Two 24V DC motors with rated speed of 5,400 rpm and gearhead ratio 6:1 were employed to provide mechanical mixing within the reactor tanks. These were connected to 2" diameter, 3-blade pitched propellers via mixing shafts that were made in-house. The diameter of the propellers was approximately one third of the reactor tank diameter.

Aeration

Aquarium-type air pumps were used to provide aeration to each reactor. Air stones were used to diffuse the supplied air. In light of the importance of the aeration step in the EBPR process, two air pumps were fitted to each reactor to provide a back-up if one should fail.

Safeguards

Float sensors were installed in the top of each SBR. If activated, these were designed to stop the LabView operation program i.e. to stop the SBRs from running. This was a feature included to prevent the SBR contents from overflowing should the tanks be overfilled. This could happen in the event of insufficient time being set for the draw phase, an effluent pump failure or a blockage in the effluent lines. Overflow of the SBRs was undesirable not only because of flooding, but also because it would cause loss of the microbial communities.

Remote inspection and operation of the SBRs was made possible via a webcam and remote access to the PC used to run them. As the reactors were designed to be operational 24 hours a day, the remote access was useful for troubleshooting problems.

For safety, electrical components were positioned on a raised shelf behind the SBR reactor tanks to reduce the likelihood of fluids coming into contact with the electrics (Figure 3-2 A).

3.3 METHODS FOR MONITORING

3.3.1 Determination of Orthophosphate

Orthophosphate concentrations were determined using an ascorbic acid colorimetric method adapted and scaled down from the standard method 4500-P E [75].

Reagent Preparation

10 mL 'combined reagent' was prepared for the analysis by mixing the following reagents: 5 mL sulphuric acid (5N), 0.5 mL potassium antimonyl tartrate (0.2743 g $K(SbO)C_4H_4O_6 \cdot {}^1/_2H_2O$ / 100 mL), 1.5 mL ammonium molybdate (4 g $(NH_4)_6Mo_7O_{24}.4H_2O$ / 100 mL) and 3 mL sodium ascorbate (0.198 g $C_6H_7NaO_6$ / 10 mL). The combined reagent was prepared fresh for each set of analyses as it was only stable for 4 hours. Ascorbic acid solution also lacked long-term stability so 10 mL was prepared once per week and stored at 4ºC.

Preparation of Calibration Curve

A stock phosphate solution (50 mg (P) / L) was prepared by dissolving 219.5 mg anhydrous KH_2PO_4 in 1 L deionised water. A working standard phosphate solution (2.5 mg (P) / L) was prepared from this stock and used subsequently to prepare a series of six standards within the phosphate range 0.15 – 1.2 mg (P) / L. To 1 mL of each of these standards as well as a blank (1 mL deionised water), 160 µL combined reagent was added and the colour was allowed to develop for between 10 and 30 minutes. The absorbance of the blank and each standard was measured at 840 nm and these values were used to construct a calibration curve. Although the standard method stipulates measuring absorbance at 880 nm, the instrument used to measure UV-Vis in our laboratory (Nanodrop 2000c, Thermo Scientific) had a maximum limit of 840 nm. The absorbance of light by molybdenum blue at 840 nm is weaker than at 880 nm but the concentration of orthophosphate in our samples measured at both wavelengths showed no significant difference between them).

Sample Preparation

Samples from the SBRs were prepared for analysis as follows: For each test, a 2 mL sample of bacterial suspension (mixed liquor) was collected and centrifuged for 5 minutes at 13 000 rpm to separate solid and aqueous phases. The supernatant from each sample was analysed for orthophosphate. Note that although the standard method stipulates using filtration

through a 0.45 µm membrane rather than centrifugation, the latter was chosen for practicality in light of the large number of samples being analysed over time. A comparison between orthophosphate determined following each separation technique yielded no significant difference in values for our samples.

Sample Analysis

To 1 mL sample supernatant, 160 µL combined reagent was added and the absorbance measured at 840 nm after between 10 and 30 minutes. The measured absorbance was compared to the calibration curve to determine orthophosphate concentration. For some samples, dilution of the supernatant with deionised water was necessary before analysis to ensure the phosphate concentration was within the calibrated, linear response range of the method. At least one phosphate standard and a blank (deionised water) was tested with each set of samples.

3.3.2 Determination of Chemical Oxygen Demand (COD)

Samples from the SBRs were analysed for Chemical Oxygen Demand (COD) as follows: For each test, 3 mL mixed liquor was sampled and centrifuged for 5 minutes at 13 000 rpm to separate solid and aqueous phases. 2 mL of the supernatant was drawn off and added to a COD test vial (Hach LCI500). The test vial was heated at 148°C for 2 hours in a dry thermostat (Hach LT200) before reading in a spectrophotometer pre-programmed to read COD values directly (Hach DR500). Note that for some samples, dilution with deionised water was necessary before analysis to ensure the COD concentration was within the specified range of the method.

3.3.3 Determination of Suspended Solids

Total Suspended Solids (TSS) and Volatile Suspended Solids (VSS) were determined using a gravimetric method based on standard methods 2540 D and 2540 E respectively [75]. For each sample, a glass-fibre filter disc (Whatman 934-AH) was placed in an aluminium weighing dish and the combined weight recorded. Subsequently, two (i.e. duplicate) 10 mL mixed liquor samples were collected from each SBR. These were collected at the end of an aerobic phase so that maximum bacterial growth for that SBR cycle had been achieved. The collected samples were filtered through the pre-weighed filter discs and the residues washed with deionised water. The filter discs were then returned to their aluminium dishes and dried at 103-105ºC for at least 4 hours. The combined disc-dish weights were recorded before

subsequently igniting the samples at 550ºC for 2 hours (including furnace heating time). The combined disc-dish weights were recorded for a final time and the TSS and VSS values (in mg/L) were calculated from the results as follows:

$$TSS = \frac{1000}{V} \times (W_{103} - W_P) \qquad \textit{Equation 3.1}$$

$$VSS = \frac{1000}{V} \times (W_{103} - W_{550}) \qquad \textit{Equation 3.2}$$

Where V = sample volume (in mL) and W_P, W_{103}, and W_{550} are the combined filter disc-aluminium dish weights (in mg) before filtration, after drying at 103-105ºC and after ignition at 550ºC respectively.

3.3.4 Measurement of Dissolved Oxygen (DO), pH and Temperature

Measurements of DO, pH and Temperature were taken with the sensors described in 3.2.2 and were logged every 5 minutes. Live readings of temperature were available with the thermocouple, however, logged readings were those taken from the pH sensor.

3.3.5 Neisser Staining for Microscopy

Light microscopy was used occasionally to examine samples of mixed liquor from each SBR. Samples were deposited directly on microscope slides and dried in air. Neisser stain was applied to the slide for 30 seconds, followed by a 1 second rinse with water. Bismark Brown – a counter stain – was then applied to the slide for 1 minute and subsequently rinsed off thoroughly. After drying in air, the slide was examined for the morphology of cells and the presence of 'Neisser-positive' (purple-stained) cells indicating the presence of polyphosphate.

3.3.6 Scanning Electron Microscopy (SEM)

Images were acquired with an FEI Quanta 200F field emission environmental SEM operated under low vacuum with a beam current of 20 keV. A backscattered electron imaging mode was used which reflects sample composition as well as topography. Samples of mixed liquor

were prepared for analysis by depositing directly on microscope slides and allowing to dry in air. No coating was applied to the samples prior to imaging.

3.4 VALIDATION OF SEQUENCING BATCH REACTORS

3.4.1 Operating Conditions

The two SBRs were seeded on 12 August 2014 with mixed liquor from the EBPR lanes (Figure 3-5) at Slough Sewage Treatment Works (Thames Water, UK) and run for an initial period of four months. Operating conditions were selected on the basis of previous studies successfully running SBRs for EBPR and are outlined below.

SBR Cycle Phase Timings

The two 2 L working-volume rectors were run on a 6-hour cycle with phase lengths based on those used by Schuler and Jenkins [76]. The length of the anaerobic and aerobic phases were 117.5 and 180 minutes respectively. The length of the fill, settle, draw and idle phases were 2.5, 56.1, 3.4 and 0.5 minutes respectively. The length of the fill phase was set according to the influent pump rate and how long it took to pump the required 1 L into the reactor tanks. The minimum 1 L volume in the reactor tanks was fixed by the position of the effluent draw pipes. The length of the draw phase was therefore set to exceed the required time for the effluent pump to draw 1 L from each of the reactors. The waste phase timing was calculated according to the target solids retention time (SRT) and the effluent pump rate. The latter was measured periodically by placing the effluent pipes into two volumetric cylinders during the draw phase and recording the fill time at each 100 mL interval up to 900 mL. Target SRTs varied between 10 and 15 days and waste phase timings varied between 4 and 7 seconds.

Figure 3-5 EBPR lanes at Slough Sewage Treatment Works, UK

Composition of the Influent – Synthetic Wastewater

The SBRs were fed with an influent composition based on that used by Schuler and Jenkins 2003 [76]. The full list of chemical components in the synthetic wastewater is summarised in Table 3-2. In particular the amount of ammonium chloride added to our system was less than that used by Schuler and Jenkins. This was due to the fact that nitrification in our reactors was inhibited by the addition of allylthiourea and hence there was less requirement for ammonia. Allylthiourea is commonly used in EBPR studies, for example [34, 77-79], to ensure no residual nitrate is present in the anaerobic phase where it would deteriorate EBPR performance. It is assumed that allylthiourea has no effect on the physiology of PAOs. To the author's knowledge, however, no study has specifically investigated this.

The carbon source used was predominantly sodium acetate which is readily biodegradable and has been used successfully in many laboratory-scale reactors. Furthermore, in full-scale systems where collected wastewater resides in sewer pipes long enough for fermentation, much of the influent organic matter is fermented to acetic acid prior to arriving at the EBPR plant [26]. Acetate is thus both an effective and relevant carbon source. COD and inorganic phosphorus concentrations were 200 and 8 mg/L respectively, yielding a COD/P ratio of 25 mgCOD/mgP. This parameter is important since phosphorus removal requires a

stoichiometric amount of COD. This stoichiometry is affected by many factors such as the nature of the carbon source, operating conditions, and the existence of organisms (GAOs) competing for COD in the anaerobic phase. It is thought by some that a COD/P ratio greater than 50 mgCOD/mgP favours the growth of GAOs over PAOs and that a lower ratio of 10-20 mgCOD/mgP should be more favourable for PAO growth [20]. Whatever the ratio used, for successful phosphorus removal, it is important that enough COD is supplied so that the process does not become COD-limited.

Table 3-2 Chemical composition of synthetic wastewater used to feed the EBPR communities.

Inorganic	117 mg/L potassium chloride (KCl); 38 mg/L ammonium chloride (NH_4Cl); 219 mg/L magnesium chloride hexahydrate ($MgCl_2.6H_2O$); 14 mg/L magnesium sulphate heptahydrate ($MgSO_4.7H_2O$); 46 mg/L calcium chloride dihydrate ($CaCl_2.2H_2O$); 40 mg/L sodium phosphate monobasic ($NaH_2PO_4.2H_2O$).
Organic	314 mg/L sodium acetate trihydrate, 30 mg/L casamino acids, 8 mg/L yeast extract, 4 mg/L allylthiourea.
Trace Element	0.061 mg/L boric acid (H_3BO_3); 0.3 mg/L zinc sulphate heptahydrate ($ZnSO_4.7H_2O$); 0.015 mg/L potassium iodide (KI); 0.061 mg/L copper sulphate pentahydrate ($CuSO_4.5H_2O$); 0.075 mg/L cobalt nitrate hexahydrate ($Co(NO_3)_2.6H_2O$); 0.031 mg/L sodium molybdate dehydrate ($Na_2MoO_4.2H_2O$); 0.34 mg/L manganese sulphate ($MnSO_4.H_2O$); 0.3 mg/L iron sulphate heptahydrate ($FeSO_4.7H_2O$).

The synthetic wastewater was prepared thus: 1 L concentrated stock solutions of the inorganic components and the trace element components were prepared whenever required and stored at 4°C. 30 L synthetic wastewater per reactor was prepared once per week and stored in fridges 4°C from which it was pumped directly into the reactor tanks. Organic components were weighed out each time a new batch of wastewater was prepared and dissolved in a minimal amount of water before adding 100 mL inorganic stock solution, 30 mL trace element solution and diluting to volume.

Hydraulic Retention Time (HRT) and Solids Retention Time (SRT)

Hydraulic retention time, HRT, is the average length of time that aqueous components remain within the SBR. It is calculated from Equation 3.3, where V_T is the volume of the

SBR reactor tank and Q_I is the influent flow rate. With an SBR tank volume of 2 L and influent flow rate of 4 L/day, the HRT was thus set at 0.5 days (i.e. 12 hours).

$$HRT = \frac{V_T}{Q_I} \qquad \textit{Equation 3.3}$$

Solids retention time, SRT, is the average time that solids (and hence individual organisms) are retained in the SBR tanks. It is calculated from Equation 3.4, where Q_W is the wasted sludge flow rate, X_W is the VSS concentration in the wasting stream, Q_E is the effluent flow rate, X_E is the VSS concentration in the effluent, $MLVSS$ is 'mixed liquor volatile suspended solids' (i.e. VSS concentration in the mixed liquor) and V_T is the volume of the SBR reactor tank. Using this equation, the volume of mixed liquor (ML) to be wasted (i.e. removed) per cycle in order to operate at a defined SRT could be calculated. Wasting was carried out at the end of each SBR cycle, just prior to settling, so that biomass growth per cycle was maximised. Solids levels (both TSS and VSS) were monitored in both reactors (generally twice per week) to check for development of a stable solids concentration – which would constitute one indicator of stable operation of the reactors. Furthermore, these measurements could be used to normalise other data such as phosphate concentration. TSS and VSS results are shown in Figure 3-6.

$$SRT = \frac{MLVSS \times V_T}{(Q_W \times X_W) + (Q_E \times X_E)} \qquad \textit{Equation 3.4}$$

For the first 10 days, no sludge was wasted while the inoculated biomass was allowed to acclimatise to new conditions within the reactor. Following this period, the SRT was set at 10 days and the solids levels (TSS and VSS) were seen to decrease steadily over time. On day 21, wasting was stopped whilst troubleshooting the cause of some unusual DO readings observed during the beginning of each anaerobic phase. Wasting was started again on day 28 after the source of the problem had been identified. This time SRT was set at 12 days – slightly longer than the initial setting in the hope that the biomass levels would stabilise. By day 44 there was no indication of the decrease in solids concentration levelling off and so wasting was stopped again in order to retain biomass in the SBRs. After a recovery period,

wasting was restarted on day 65 with the same SRT (12 days). Eventually, toward the end of the validation period, a levelling-off in TSS seemed to occur at concentrations of around 1000 and 700 mg/L respectively for SBR-A and SBR-B. These are fairly low concentrations so it was decided to increase the SRT to 15 days in the next phase of SBR operation.

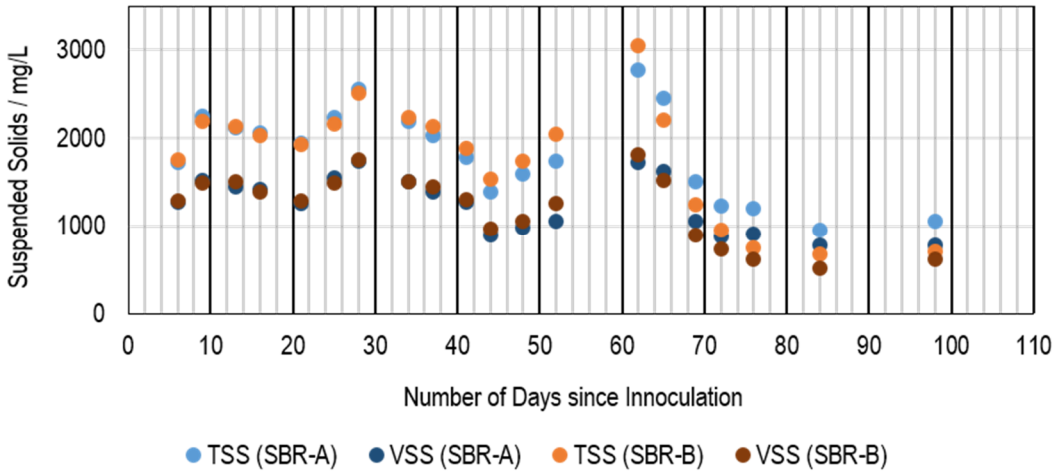


Figure 3-6 TSS and VSS levels in SBR-A and SBR-B during the validation phase.

Temperature, pH and Dissolved Oxygen

Temperature and pH were not controlled parameters in the systems built here. Limited control of DO levels was possible by altering air flow rate (with a crude, uncalibrated valve) and aeration timings. Values for temperature, pH and DO were monitored and logged at 5 minute intervals. Data points of interest extracted from each SBR cycle are presented in Figure 3-7.

Investigations into the effect of temperature on EBPR performance over past decades have yielded contradictory findings. Between temperatures of approximately 5 and 24°C, some have reported no effect of temperature on EBPR performance whilst others have reported that low temperatures affect performance adversely [32]. At higher temperatures (30°C) it is generally understood that GAOs can assimilate acetate faster than PAOs and so phosphorus removal performance can deteriorate due to a population shift toward GAOs [80].

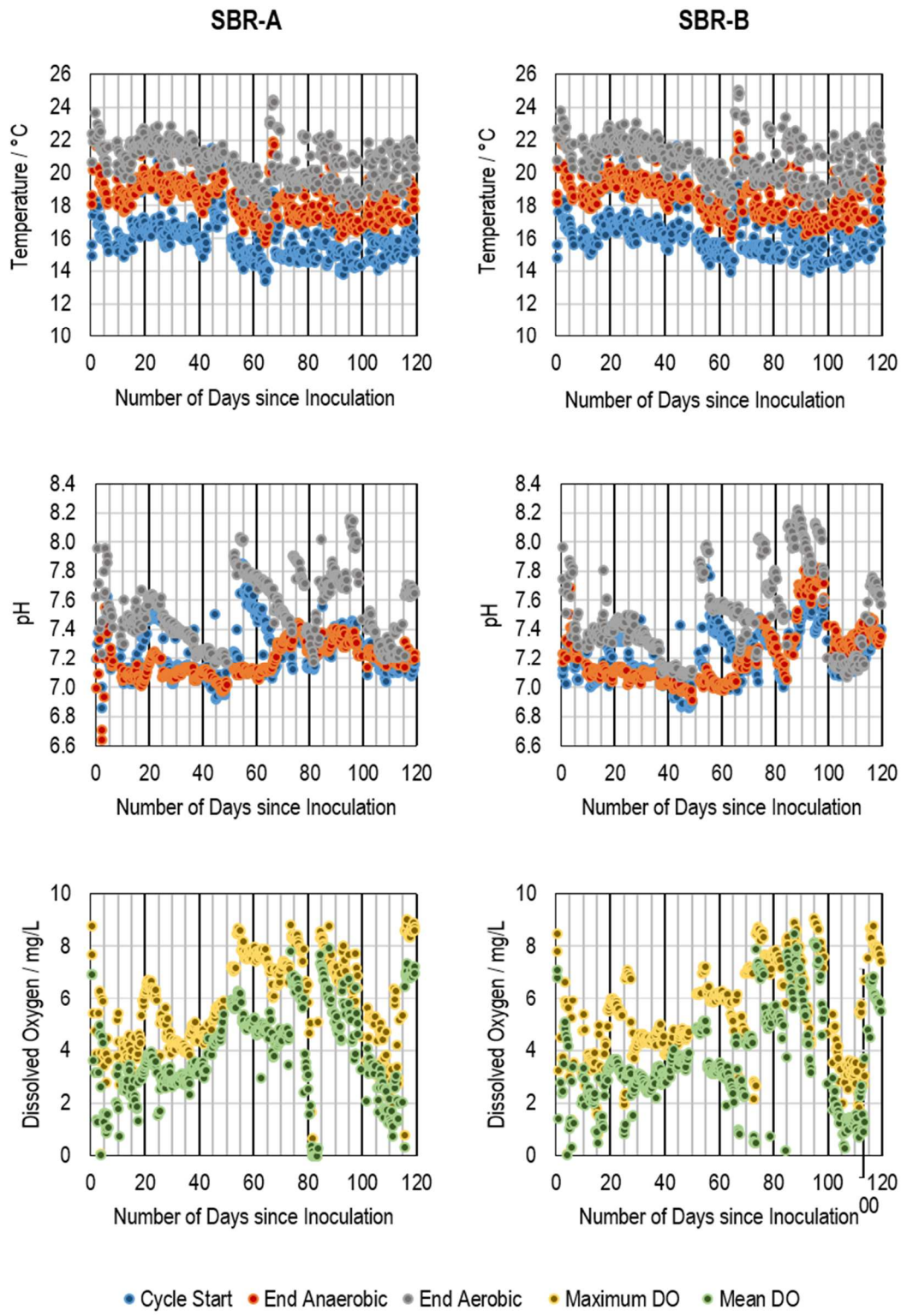


Figure 3-7 Temperature (top), pH (middle) and Dissolved Oxygen (bottom) for SBR-A (left) and SBR-B (right) during the validation phase. Dissolved Oxygen is abbreviated as DO in the legend.

It can be seen in Figure 3-7 that the temperatures in the reactors fluctuated between an average of 16.0°C (SD=1.2) at the beginning of each cycle when the new feed was pumped in, and an average of 20.7°C (SD=1.2) at the end of each cycle depending on the ambient air temperature. The cyclic fluctuation was due to the fact that synthetic wastewater was prepared once per week and stored at 4°C to limit bacterial growth and thus preserve COD levels. Pumping the refrigerated influent into the reactors at the start of each cycle thus lowered temperature in the reactor tanks considerably. On day 65, a gentle heat mat (originally purposed for seed germination) was installed under the reactors because of falling ambient room temperature with the seasonal weather. A thermostat was used so the heat mat was switched on only when ambient air temperature dropped below 18°C. If temperatures had been controlled, the target temperature would have been 20°C so that it would be in the range understood to aid uptake of acetate by PAOs, as well not be too cold as to slow bacterial growth and metabolism. What was actually observed was that in the anaerobic phase, when the competition of PAOs over GAOs to take up acetate is critical, the temperatures were cooler – ranging from a minimum of 13.4 °C at the start of a cycle to a maximum of 22.3°C at the end of anaerobic phase as the reactors warmed up to room temperature. During the aerobic phase, where the assimilated carbon compounds are metabolised and bacterial growth occurred, the temperatures were warmer – ranging between a minimum of 17.4 and a maximum of 25 °C.

Full-scale wastewater treatment plants (WWTPs) operating EBPR tend to be continuous flow rather than batch processes. Temperature in such plants is largely a result of the local climate since it tends to be monitored but not controlled. Seasonal and diurnal temperature variation would be expected albeit buffered against fast fluctuations by the greater volume of wastewater with respect to the bench-scale lab reactors. In 2015, mean temperatures across the UK were recorded by the Met Office to be 3.9, 7.7, 13.9 and 10.0 °C during winter, spring, summer and autumn respectively [81]. These temperatures are cooler than the operating temperature of the SBRs in this work. In addition to climate effects, however, thermal inputs from chemical reactions and mechanical operation of the plants will have an impact on temperature. Jaffer *et al.* reported wastewater temperatures at different locations across the entire Slough WWTP (Thames Water, UK) to range from 13 to 26.8 °C [18]. The coolest temperature (13 °C) was recorded closest to the EBPR plant within the works and is thus the most comparable to the lab-scale experiments run in this work. The warmest temperature (26.8 °C) was recorded in the digester. It was not stated at what time of year these measurements were made.

The optimum pH range for EBPR performance is thought to lie between 7 and 8. Filipe *et al.* observed in a lab-scale SBR, that the rate of acetate uptake by PAOs was independent of pH between 6.5 and 8. For GAOs, however, acetate uptake rate decreased with increasing pH range. At pH 7.25, acetate uptake by the two populations (PAOs and GAOs) occurred at an equal rate. Above this threshold PAOs could assimilate acetate faster than GAOs [31]. Zhang *et al.* observed that when the pH in a lab-scale SBR was lowered from 7.0 to 6.5, phosphorus removal performance dropped significantly from 99.9% at pH 7.0 to 17% at pH 6.5 [82]. Above pH 8, acetate uptake rate by PAOs has been observed to decrease [83].

It can be seen in Figure 3-7 that the majority of pH data collected over the 4 months validation phase was between the preferred range of 7 and 8. In addition, it was noted that pH recorded in the reactors at the end of the anaerobic phase were generally lower than those recorded at the end of the aerobic phase. The trend observed in pH recorded at the end of the aerobic phase seemed to resemble the trend in DO concentration measured during the aerobic phase. Figure 3-8 shows that there was indeed a moderate correlation between these two parameters. This may be due to the fact that aeration increases the area of the mixed liquor-air interface and thus facilitates decarboxylation of the mixed liquor. Since aqueous carbon dioxide is acidic, transfer of carbon dioxide from the mixed liquor to air causes pH in the reactors to increase.

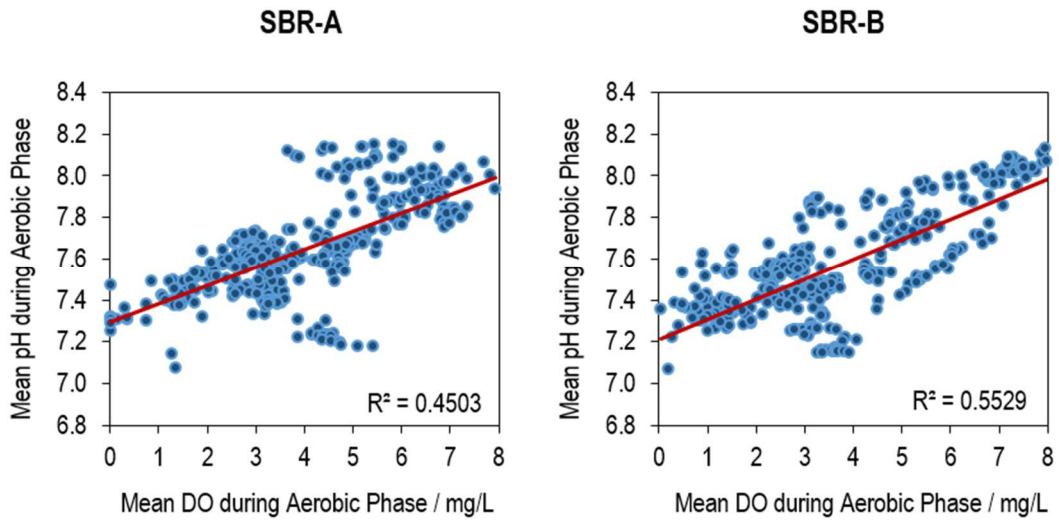


Figure 3-8 Paired plots of mean pH and mean dissolved oxygen (DO) measured during the aerobic phase for SBR-A (left) and SBR-B (right).

Very few studies have been carried out to investigate the effect of dissolved oxygen concentration on EBPR performance. The 'integrity' of the anaerobic phase is known to be important – if oxygen is present in this phase then PAOs lose some of their competitive edge over other organisms and furthermore, PHA stores within these organisms can be gradually depleted leaving them unable to re-assimilate phosphate in the aerobic phase. As a result, EBPR performance is adversely affected [84]. With respect to the aerobic phase, across several full-scale EBPR plants, Griffiths *et al.* more frequently observed poor phosphate removal performance at relatively high DO concentrations (4.5-5.0 mg/L). They also observed a greater number of PAOs in plants operated with DO levels between 2.5 and 3.0 mg/L [85]. This seems to suggest that although aeration is vital for biological wastewater treatment, excessive aeration may be detrimental to process performance. This is good news for the industry since aeration constitutes one of the major costs of the process. Further studies are required, however, to confirm these findings.

In this study, a target DO concentration of between 3 and 5 mg/L was selected for the aerobic phase. Without a feedback system, however, dissolved oxygen levels were incredibly difficult to regulate. This is reflected in the monitoring data (Figure 3-7) which shows the maximum and mean DO per cycle aeration phase. With relatively fast air flow and/or constant aeration, dissolved oxygen levels would exceed 8 mg/L. If the air flow was restricted too much, however, the DO concentration in the reactor tanks would take a long time, sometimes as long as an hour, to reach the target level. This was presumably because at the beginning of the aeration phase bacterial metabolism occurred at a faster rate than later in the phase and so any available oxygen was used up quickly. In light of this, another strategy that was attempted was to set one aerator to provide a continuous low level oxygen supply, and to use the other aerator to act as a high level short-term (about 15 minutes) 'boost' oxygen supply at the beginning of the aerobic phase. This was the strategy that was eventually employed and was achieved (until an opportunity to update the LabView programme is available) by installing plug-in timers between the aerators and the control interface. These timers were also used to switch aeration off 30 minutes before the end of the aeration phase to allow the residual DO to be used up before the start of the next cycle.

The large variation in maximum and mean dissolved oxygen concentration shown in Figure 3-7 indicates varying success at keeping the DO levels in the target range with the existing set-up. Alongside the difficulty of controlling air flow itself, variation in both solids (i.e. biomass) levels and temperature also affects DO concentration. As the biomass level increases, DO is consumed faster, and as temperature increases the solubility of DO

decreases. In light of all these factors, the only effective way to control DO concentrations at a target level would be to introduce a feedback system to switch the aerators on and off when required according to the readings logged by the DO sensors. In addition, introducing temperature control would contribute significantly to improving stability in both biomass and DO concentrations.

3.4.2 Performance

The presence of PAOs in each reactor was indicated by microscopic examination of sludge samples. Clusters of cocci-shaped organisms were seen to stain 'Neisser-positive' (purple) for the presence of polyphosphate Figure 3-9.

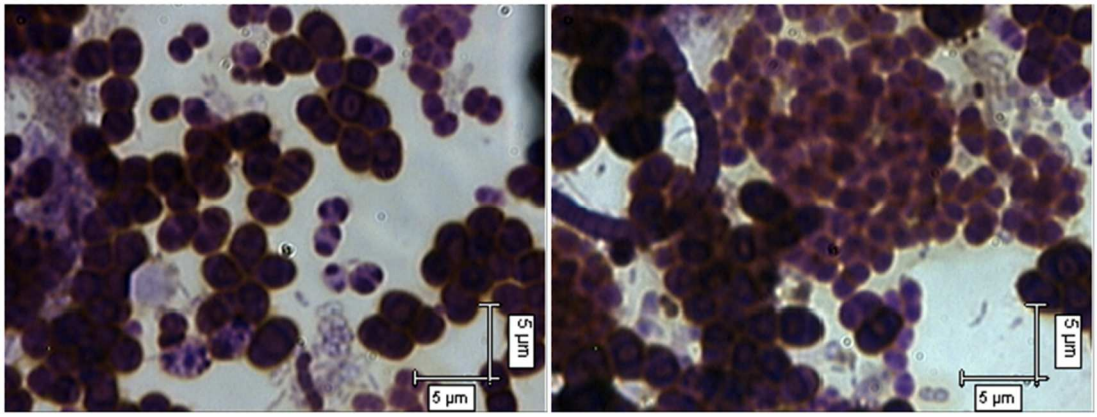


Figure 3-9 Micrographs of mixed liquor from SBR-A (left) and SBR-B (right) sampled on day 18 after inoculation. Samples were stained with Neisser stain and counterstained with Bismark Brown. A 100× objective gave a sample magnification of 1000× and scale bars indicate a length of 5 µm.

Characteristic EBPR behaviour is indicated by aqueous orthophosphate concentration rising toward a maximum at the end of the anaerobic phase as intracellular polyphosphate is broken down and released in exchange for short-chain carbon compounds. Subsequently aqueous phosphate concentration diminishes toward a minimum at the end of the aerobic phase as phosphate is taken up from solution to replenish polyphosphate stores. Successful EBPR performance is defined by consistent achievement of orthophosphate concentration below the regulated phosphorus discharge consent at the end of the aerobic phase. For this work a target limit of 1.0 mg/L PO_4-P in the effluent was defined in accordance with current

European regulations for discharge of wastewater to sensitive areas which are subject to eutrophication [86].

Phosphate removal (i.e. EBPR) performance in SBR-A and SBR-B is shown in Figure 3-10. It can be seen that SBR-A displayed characteristic EBPR behaviour and reduced the influent phosphate to zero throughout most of the validation phase. There were periods of time within this, however, where the concentration of phosphate measured at the end of the anaerobic phase showed a downward trend with time. These downward trends were observed between day 28 and day 48, and between day 52 and day 76 and are likely explained by the concurrent decrease in suspended solids levels. On day 48, no phosphate was released during the anaerobic phase, and the effluent phosphate concentration of 2.4 mg/L exceeded the phosphorus limit. The reason for this is not known. By day 52, however, the reactor had recovered characteristic EBPR behaviour.

The phosphate concentration trends in SBR-B mostly mirrored those of SBR-A in successfully removing phosphorus from the synthetic wastewater. From day 84 until the end of the validation phase, however, effluent phosphorus concentrations exceeded the target limit of 1.0 mg/L and indicated that EBPR performance in the reactor had been lost almost entirely. The reason for this is not known, but perhaps could have been triggered by an aeration failure in 2 successive cycles that occurred on day 87 due to a pump failure (Figure 3-7).

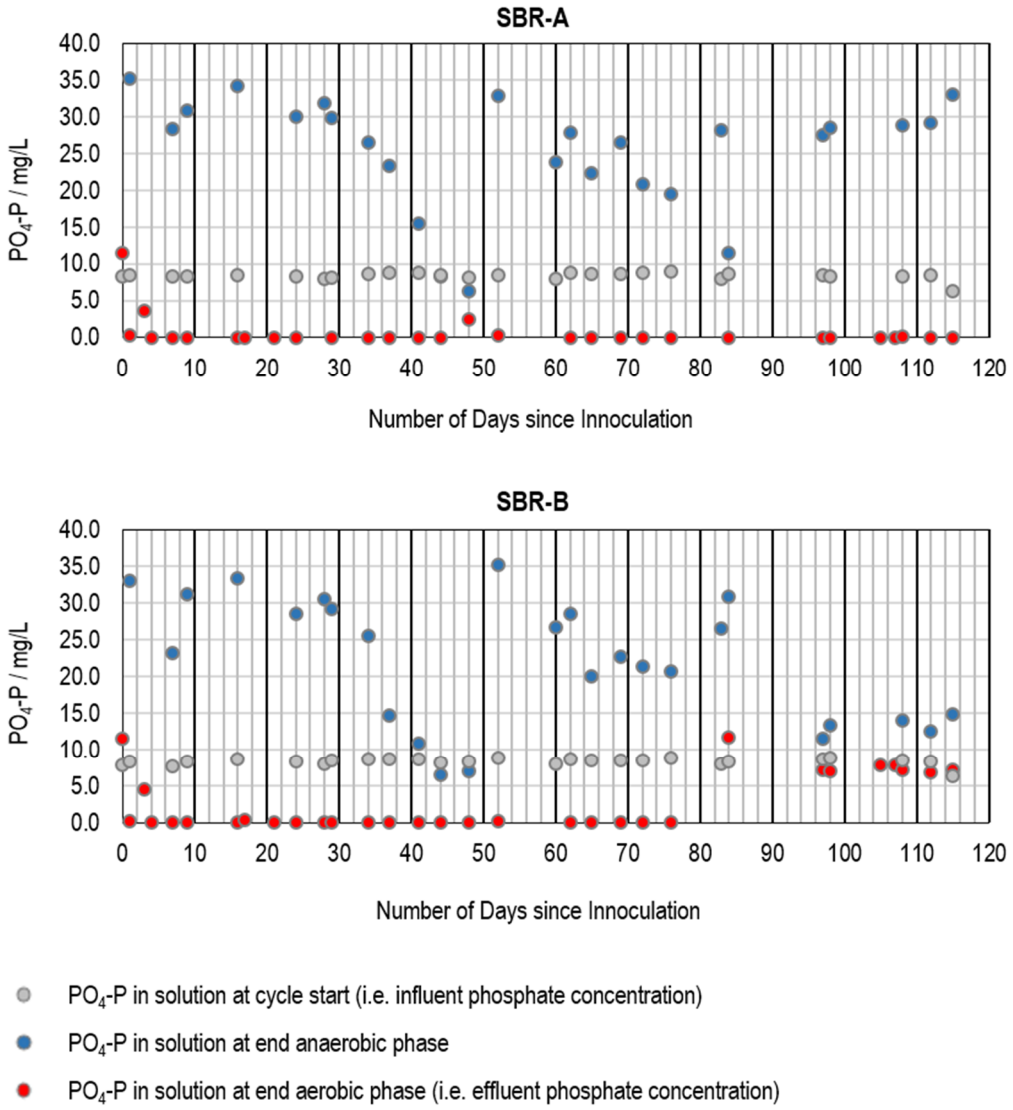


Figure 3-10 PO_4-P removal performance of SBR-A (top) and SBR-B (bottom) measured during individual anaerobic-aerobic cycles during the validation phase.

3.4.3 Second Run of the SBRs

The second run of the SBRs commenced immediately after the end of the validation phase. The two reactors were seeded with mixed liquor collected from Slough Sewage Treatment Works (Thames Water, UK), rather than using frozen stocks prepared from SBR-A and SBR-B before shutdown. The two SBRs were renamed SBR-C and SBR-D during this phase to avoid any confusion in data reporting. Operating conditions and performance of the reactors for 10 months of this phase are briefly described below.

The SBRs were run with the same SBR cycle times as for the validation phase and were fed with the same synthetic wastewater composition. The target SRT was set to 15 days and the measured solids content in the reactors are shown in Figure 3-11.

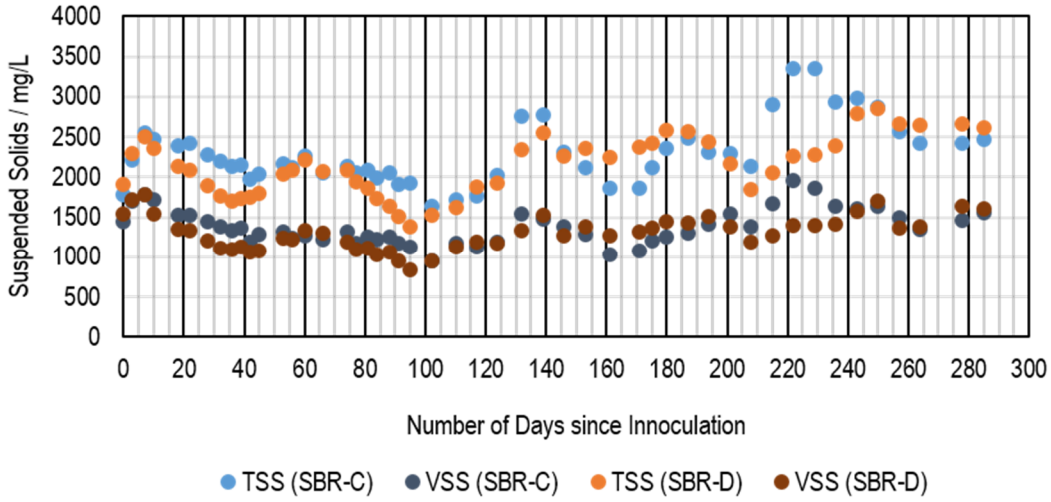


Figure 3-11 TSS and VSS levels in SBR-C and SBR-D during the second run of the SBRs.

Monitoring data for temperature, pH and DO are shown in Figure 3-12. Temperatures in the reactors fluctuated between an average of 15.8 (SD=1.2) at the beginning of each cycle, and an average of 20.5 (SD=1.1) at the end of each cycle, with the exact temperatures depending on the ambient air temperature. pH levels initially varied between approximately 7.0 and 7.6 throughout a cycle, however, over time, the pH levels showed a general downward trend. By the end of 10 months of operation, the pH levels in the reactors varied between approximately 6.8 and 7.2 over the course of a cycle. According to the previous studies already discussed, this may have caused a population shift away from PAOs and toward GAOs which are thought to be able to assimilate carbon faster than PAOs at pH levels below 7.25. No adverse effect on the phosphorus removal performance of the reactors, however, was observed to be concurrent with these lower operational pH values. With respect to dissolved oxygen levels, once again, the large variation in maximum and mean DO concentrations demonstrated little success at keeping this parameter within the target range with the current system.

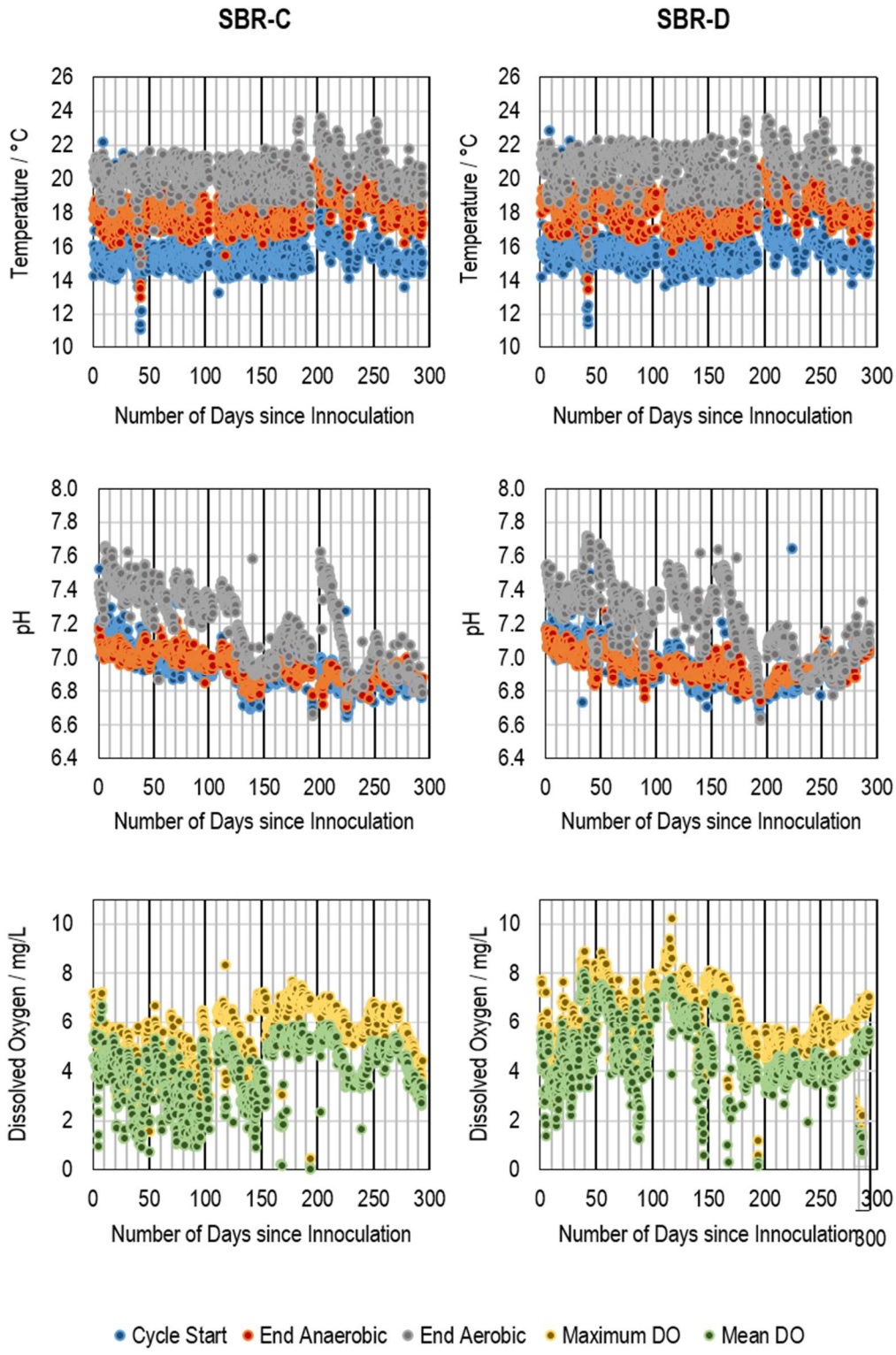


Figure 3-12 Temperature (top), pH (middle) and Dissolved Oxygen (bottom) for SBR-C (left) and SBR-D (right) during the second run of the SBRs. Dissolved Oxygen is abbreviated as DO in the legend.

Once again, presence of PAOs in each reactor was indicated by microscopic examination of sludge samples and clusters of cocci-shaped organisms were seen to stain 'Neisser-positive' (purple) for the presence of polyphosphate (Figure 3-13). Furthermore, SEM imaging of dried sludge samples deposited on slides showed the presence of clusters of light-coloured spheres. As the imaging was conducted in backscattering mode, the composition as well as the morphology of the sample is represented by the image. Energy-dispersive X-ray spectroscopy confirmed that phosphorus and oxygen were principal elements contained in these clusters and thus they were assumed to be polyphosphate granules contained within clusters of EBPR bacteria.

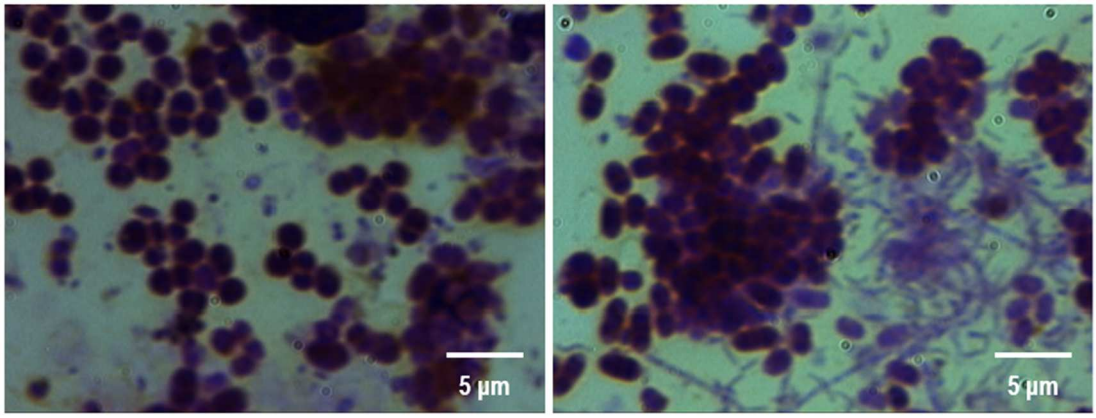


Figure 3-13 Micrographs of mixed liquor from SBR-C (left) and SBR-D (right) sampled on day 53 after inoculation. Samples were stained with Neisser stain and counterstained with Bismark Brown. A 100× objective gave a sample magnification of 1000× and scale bars indicate a length of 5 µm.

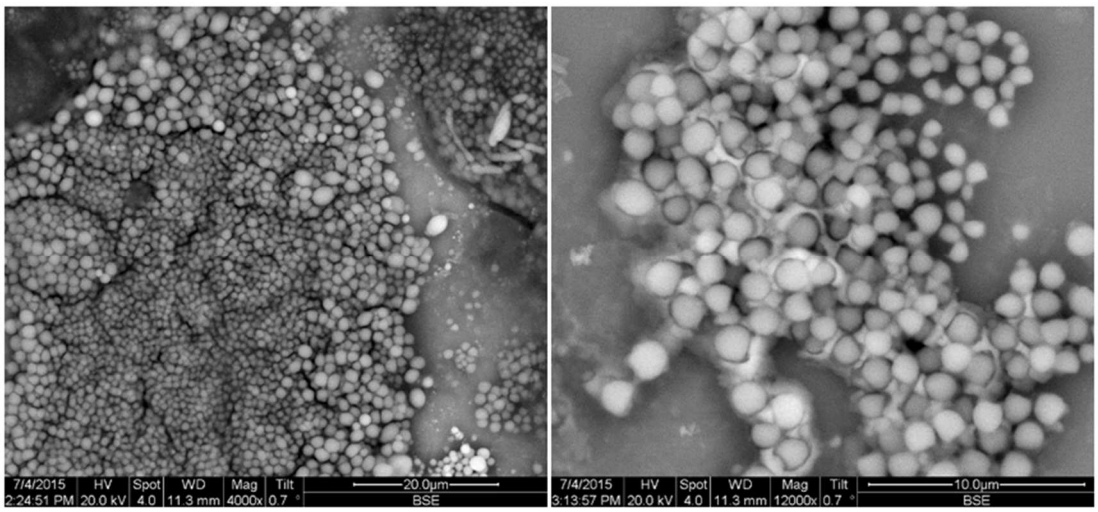


Figure 3-14 SEM micrographs of mixed liquor sampled from SBR-D at 4000× (left) and 12000× (right) magnification. Scale bars indication a length of 20 µm (left) and 10µm (right).

Phosphate removal (i.e. EBPR) performance in SBR-C and SBR-D is shown in Figure 3-15. Both SBRs displayed characteristic EBPR behaviour with the exception of one sampling point in SBR-D on day 53 when phosphate concentration in the mixed liquor did not increase by the end of the anaerobic phase. This is perhaps correlated to a disturbance recorded in SBR operation where, immediately prior to this measurement, a power cut occurring during the weekend had caused the reactors to sit idle for 2 days. Nonetheless, characteristic EBPR behaviour was subsequently restored in the reactors. Phosphorus concentrations measured in the effluent were consistently below the target 1.0 mg/L limit during the entire period of operation. The performance of these EBPR reactors can thus be considered to be successful.

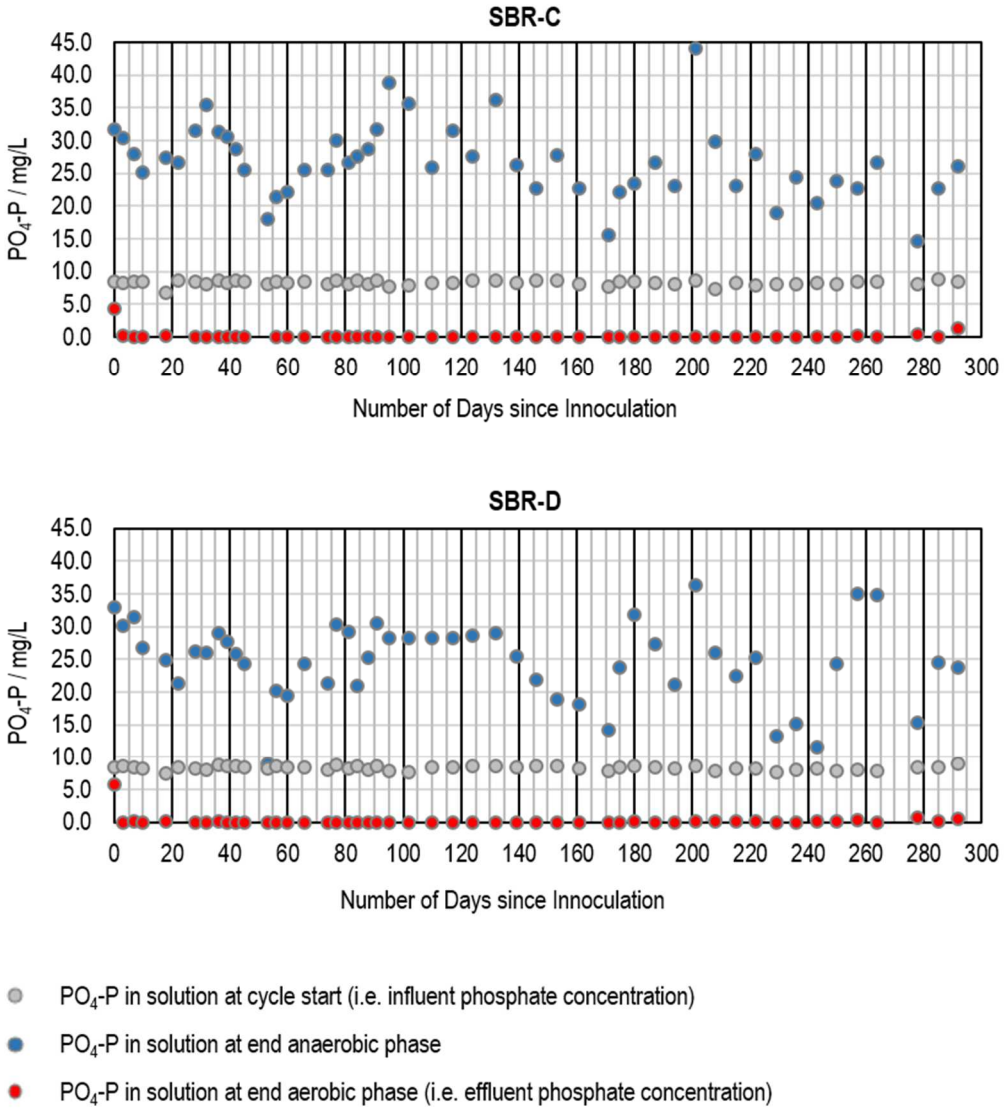


Figure 3-15 PO_4-P removal performance of SBR-C (top) and SBR-D (bottom) measured during individual anaerobic-aerobic cycles during the second run of the SBRs.

3.4.4 Summary of Modifications made during SBR Operation

A number of modifications were made to the sequencing batch reactors during operation, some of which have already been discussed. For convenience, all modifications made during either the validation or second run of the SBRs are summarised below.

The target SRT in the reactors was increased from the trialled 10 days, to 12 days, then up to 15 days for the second SBR phase. This was done in response to the observation that at an SRT of 10 days, the solids levels in the SBRs failed to stabilise at a desirable concentration.

The aeration regime in the tanks was altered from having a continuous moderate-rate supply to using one aerator to provide a continuous low-level supply and using the other aerator to provide a high-rate 'boost' supply for 15 minutes at the beginning of the aerobic phase. This was achieved with plug-in timers installed between the control interface and the aerators. Furthermore, these timers were used to switch off the air supply 30mins before the end of the aerobic phase to allow all the available DO to be used up before the start of the subsequent SBR cycle. This was to preserve the integrity of the anaerobic phase.

Control of the air flow from each of two aerators was changed from individual screw-type valves to two 4-way, 4 mm diameter manifolds with lever valves. This was initially because two of the aeration pumps stopped functioning within a short time of each other. Restricting the air flow quite severely using the manufacturer supplied valves created a lot of back pressure on the pumps. It was hypothesised that this may have shortened the lifespan of the aerators. The manifolds enabled air flow into the reactor tanks to be reduced by bleeding off some of the supply rather than restricting its flow and creating back pressure. Since installing the manifolds, no aerator has failed. The other added benefit of the manifolds was that air supply from one aerator could be split between the two reactor tanks.

One-way check valves were installed in each effluent line between the reactor tank and the effluent pump on day 137 of the second run. This modification was made after it was observed that the sludge blanket in the reactors was being disturbed and re-suspended during the draw phase. This meant that more sludge than intended was being wasted. The occurrence of this was linked with the effluent pumping process, because the location of disturbed sludge within both tanks was directly underneath the draw pipe. It was assumed that some kind of backflow or slippage could be occurring during pumping. To fix the problem and prevent it happening again in future, check valves were installed in the effluent lines. In future SBR operation, it is worth investigating if increasing the occlusion setting on

the pump heads could also remedy this problem (although at higher occlusion settings, pump tubing is known to have a shorter lifespan).

During the colder months a heat mat was placed under the reactors to help the reactors approach a temperature of 20°C by the end of the aerobic phase. The thermostat was set at 18°C so that when room temperature dropped below this, the device was switched on.

Although perhaps trivial, it proved effective that after making two errors in operation by leaving valves turned to the wrong position or leaving manual override switches on, an end-of-day checklist was installed by the reactors. As a result of this, no further errors were made in SBR operation.

3.5 FUTURE IMPROVEMENTS FOR SEQUENCING BATCH REACTORS

Based on experience with SBRs A, B, C and D, recommendations for future improvements to the current system include the following.

Programming Changes

Programming changes to LabView should be made to integrate sensor data for DO with control of the aerators so that aeration can be operated using a feedback loop to ensure DO during the aeration phase is kept within a target range. Integrating the sensors into the LabView programme would have the added benefit of allowing a distinction between the user-defined time interval for logging readings and the time between 'live' display readings. Another proposed change to the LabView SBR program is to add an extra cycle phase between the aerobic and settling phase. This phase would comprise mixing but no aeration in order to use up any residual DO in the reactors before the start of the subsequent cycle. This aids in preserving the integrity of the anaerobic phase to ensure a selective advantage is given to PAOs.

Improved Control of Influent and Effluent Volumes

The current method for controlling the influent and effluent volumes pumped in and out of the SBR tanks is fairly crude, being based upon peristaltic pump rates and time. Variation in the pump rates can occur over time as the pump tubing wears. Furthermore, one pump is shared for the influent and another pump is shared for the effluent. This is economical but the pump rate for both heads is never equal. If more accurate volume control is required,

therefore, separate pumps should be employed for separate reactors and level sensors in the reactor tanks should be installed to instruct these pumps when to switch on and off.

Temperature Control

Introducing temperature control to the system is recommended to aid stability of parameters such as biosolids levels and DO concentration.

Sensor Issues

The Ponsel sensor readings for DO and pH were often found to go 'offline' and stop measuring the required parameters until the monitoring program was rebooted. One activity observed to cause this was switching nearby mains electricity supplies on or off so perhaps the sensors are particularly sensitive to fluctuations in the electricity supply. In any case, further investigation of the cause and potential solution to this problem is warranted.

In Situ Raman Probes

With the current SBRs, EBPR mixed liquor was sampled, dried on CaF_2 slides, and analysed off-line with the available Raman spectrometer. A significant next step for this work would be to incorporate fibre optic Raman probes into the SBRs to obtain measurements of live EBPR bacteria in situ.

Chapter 4
Methods II: Raman Spectroscopy of Individual Bacteria

4.1 INTRODUCTION TO RAMAN SPECTROSCOPY 57
4.1.1 Molecular Vibration 57
4.1.2 The Raman Effect 59
4.1.3 The Raman Spectrum 62
4.1.4 Confocal Raman Micro-Spectroscopy (Raman Microscopy) 65
4.2 ACQUISITION OF RAMAN SPECTRA 66
4.2.1 Instrumentation 66
4.2.2 Spectral Resolution 68
4.2.3 Spatial Resolution 70
4.2.4 Experimental Estimation of Raman Sampling Volume 71
4.2.5 Instrument Alignment 78
4.2.6 Acquisition Parameters 78
4.2.7 Sample Preparation 80
4.3 SPECTRAL PREPROCESSING 81
4.3.1 Cosmic Spike Removal 81
4.3.2 Background Removal and Baselining 81
4.4 OVERVIEW OF SPECTRAL ANALYSIS 82
4.4.1 General Methods 82
4.4.2 Multivariate Methods 82

4.1 INTRODUCTION TO RAMAN SPECTROSCOPY

Raman, infrared (IR) and near-infrared (NIR) spectroscopies are all types of molecular vibrational spectroscopy. They use the interaction of light and the characteristic vibrational modes of a material to yield physicochemical information about that material. Whereas IR and NIR spectroscopy rely on the absorbance of light to probe vibrational energy levels, Raman spectroscopy relies on the inelastic scattering of light [87]. A brief introduction to the theory of Raman spectroscopy is given here, preceded by a short description of molecular vibration. More comprehensive and mathematical descriptions of these concepts can be found in literature [88-90].

4.1.1 Molecular Vibration

The constituent atomic nuclei within a molecule are not fixed in space relative to each other. Rather, the chemical bonds between these nuclei stretch and bend in periodic motion giving rise to molecular vibrations. Typically, a molecule with N atoms has $3N-6$ 'normal vibrational modes' (henceforth abbreviated to 'vibrational modes') that arise from $3N$ degrees of freedom. The deducted modes correspond to six modes that result in translation and rotation of the molecule rather than vibration. Note that a linear molecule has $3N-5$ vibrational modes since only two rotational modes together with the three translational modes can be observed. Each vibrational mode represents a specific vibrational pattern within the molecule for which all atoms vibrate at the same frequency. This pattern is described by a linear combination of Cartesian coordinates of the nuclei that is known as the normal coordinate, Q. Normal coordinates are used to describe vibrational amplitudes for each normal mode.

For a particular vibrational mode to be Raman 'active', it must give rise to a change in polarisability, α, of the molecule. In mathematical terms, this selection rule is expressed as

$$\left(\frac{\partial \alpha}{\partial Q}\right)_0 \neq 0 \qquad \textit{Equation 4.1}$$

where Q is the normal coordinate and hence represents the amplitude of the vibrational mode, and the subscript zero denotes the equilibrium position of the vibrational mode.

Polarisability in this context is the ease with which the shape and distribution of the electron cloud may be distorted in an electric field. The effect of a molecular vibration on polarisability is difficult to predict, but an analogy with the volume of the electron cloud may be used to help make qualitative predictions [91]. This works because a change in the volume of the electron cloud will inevitably affect its polarisability. For example, in Figure 4-1, symmetric stretching of the linear carbon dioxide molecule increases the volume of the electron cloud, thereby changing the polarisability of the molecule. This vibrational mode is therefore Raman active. On the other hand, the asymmetric stretching and bending modes shown have little effect, if any, on the volume of the electron cloud. In addition, for these vibrational modes the polarisability is symmetrical around the equilibrium position. The derivative of polarisability with respect to the normal coordinate is therefore zero at the equilibrium position and the vibrational mode is Raman inactive.

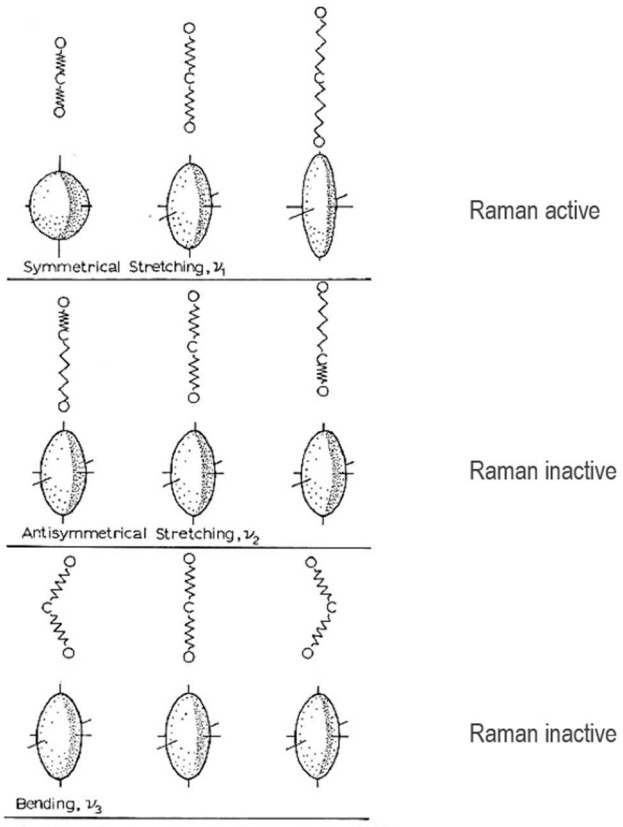


Figure 4-1 Change in polarisability and hence Raman 'activity' with the molecular vibrational modes of a linear triatomic molecule. Taken from [92].

Some vibrational modes – where the bonds of a functional group stretch and bend significantly while the remaining nuclei remain relatively still – tend to be characteristic of functional groups. Other vibrational modes are much more complex involving the relative motion of many nuclei thus being more characteristic of the molecule as a whole. The vibrational frequencies of each mode are related to the bond strengths and masses of the nuclei involved. The combination of vibrational modes present and their frequencies are thus characteristic for any particular molecule, giving rise to unique Raman spectra.

4.1.2 The Raman Effect

Modern-day Raman spectroscopy involves irradiating a sample with a laser and detecting the frequency and intensity of the scattered light. The interaction between light and the sample that gives rise to frequency-shifted scattered light is known as the Raman effect. This is described below in terms of both classical and quantum mechanical theory.

Classical Description

If light is considered as an electromagnetic wave, the Raman effect may be described as follows. The electric field component of light incident on a molecule induces charge polarisation i.e. an electric dipole moment in that molecule. In simplified terms, this induced dipole moment, μ, is directly proportional to the external electric field, E, with the constant of proportionality being the polarisability of the electron cloud, α:

$$\mu = \alpha E \qquad \textit{Equation 4.2}$$

Since the incident electric field oscillates as a sine wave the induced dipole also oscillates and in-turn generates secondary electromagnetic waves.

At any particular moment in time, t, the induced dipole moment, μ, is described by the equation:

$$\mu = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \frac{1}{2}\left(\frac{\partial\alpha}{\partial Q_m}\right)_0 Q_{m0} E_0 \times$$

$$\{\cos[2\pi(\nu_0 + \nu_m)t\,] + \cos[2\pi(\nu_0 - \nu_m)t]\}$$

Equation 4.3

where α_0 is the polarisability of the molecule at equilibrium nuclear configuration, and E_0 and ν_0 are the amplitude and frequency of the electric field component of the incident radiation respectively. Q_{m0} and ν_m are the vibrational amplitude and frequency respectively for vibrational mode m of a molecule.

The first term in this equation describes Rayleigh scattering – where the secondary radiation generated by the induced dipole is observed at the same frequency, ν_0, as the incident radiation. The second and third terms in this equation describe secondary radiation which has been modulated by the frequency of a vibrational mode, m. These 'beat' frequencies, $\nu_0 \pm \nu_m$ are what constitute Raman-scattered light (Figure 4-2).

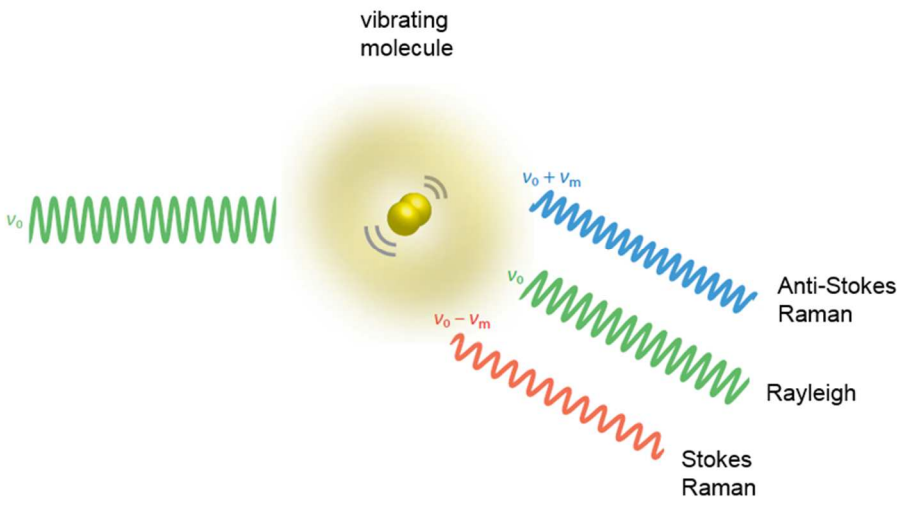


Figure 4-2 Schematic representation of Rayleigh (elastic) and Raman (inelastic) scattering of light. The Raman scattered light is modulated by the frequency of molecular vibration, ν_m. Taken from [91].

Quantum Mechanical Description

Light may be considered as discrete units i.e. photons, each with an energy E,

$$E = h\nu = hc\bar{\nu} \qquad \textit{Equation 4.4}$$

where ν is the frequency of the light, $\bar{\nu}$ is the wavenumber, h is Planck's constant and c is the speed of light in a vacuum.

When a photon encounters a molecule it may transfer energy so that the molecule is excited to a virtual higher-energy state. This state is short-lived, and the molecule quickly relaxes to a stable energy level with simultaneous emission of a photon. If the molecule returns to the initial vibrational energy state, Rayleigh (elastic) scattering has occurred and the emitted photon has the same energy as the incident photon. If, however, the molecule returns to a different vibrational energy level, then Raman (inelastic) scattering has occurred: there has been a quantised energy transfer between the incident photon and the molecule, and the energy (i.e. frequency) of the emitted photon may be used to gain information about the characteristic vibrational energy levels of the molecule (Figure 4-3).

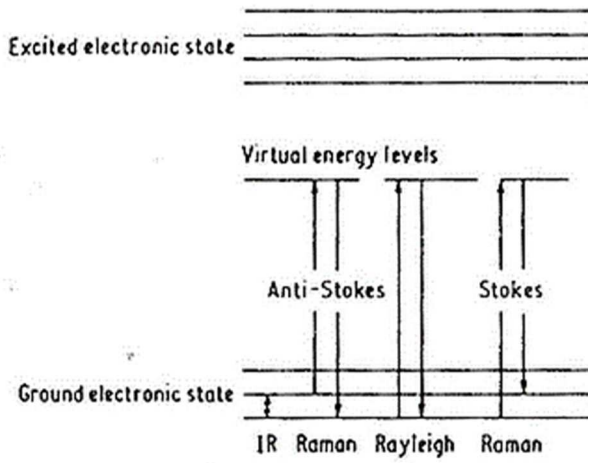


Figure 4-3 Diagrammatic representation of molecular energy levels and transitions corresponding to Infrared, Rayleigh and Raman scattering. Adapted from [93].

Frequency of Scattering Events

Raman scattering is a rare event. Whereas approximately one in 10^3 to 10^4 incident photons may be Rayleigh scattered, the number that are Raman scattered are 3-5 orders less than this [88]. At ambient temperature, a Boltzmann distribution of energy prescribes that most Raman scattered light will be Stokes (red-shifted) rather than anti-Stokes (blue-shifted). For this reason, Raman spectra taken at ambient temperature typically display the more intense Stokes Raman frequencies.

4.1.3 The Raman Spectrum

A Raman spectrum is a plot that represents the position (frequency) and intensity (number of events) of Raman scattering. By convention, the 'Raman shift' (i.e. the difference between incident and Raman-scattered light frequencies) is plotted rather than absolute frequencies. Raman shift, $\bar{\nu}$, is given by:

$$\bar{\nu} = \frac{1}{\lambda_0} \pm \frac{1}{\lambda_s} \qquad \textit{Equation 4.5}$$

where λ_0 is the incident light frequency and λ_s is the scattered light frequency.

A 'full' Raman spectrum of carbon tetrachloride depicting both absolute frequencies and Raman shifts is shown in Figure 4-4. The term 'full' here refers to the fact that Raman spectra typically show only the peaks arising from Stokes Raman scattered light since this encompasses the most useful information.

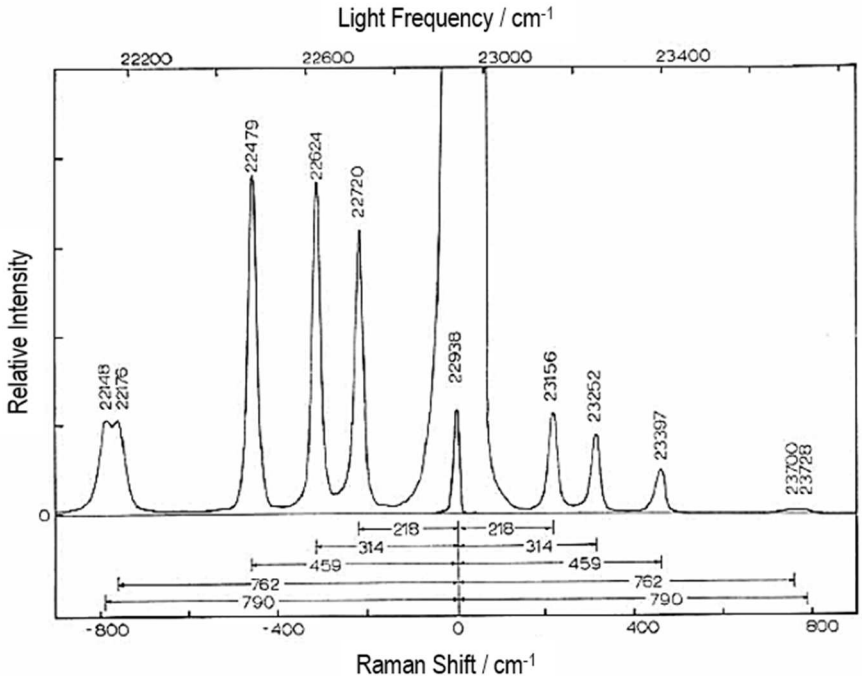


Figure 4-4 Example Raman spectrum of the tetrahedral molecule carbon tetrachloride (CCl_4), excited with 436 nm light from a mercury arc. Reproduced from [92].

The sample in this figure was irradiated with light from a mercury arc (λ_0 = 436 nm). Expressed as wavenumbers this is 22938 cm^{-1} and the most intense scattering line at the centre of the spectrum is due to Rayleigh scattering which does not shift this frequency. To the left and right of this, the bands arising from Stokes and anti-Stokes Raman scattering are shown respectively. The Stokes peaks are stronger than the anti-Stokes peaks since the former is a more common scattering event than the latter. The Stokes peaks are negatively shifted (corresponding to a loss in vibrational energy) and the anti-Stokes peaks are positively shifted with respect to the incident radiation frequency. Despite this, for simplicity most Raman spectra plot a relative wavenumber scale [88] where:

$$\bar{\nu} = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_s}\right) \times \frac{10^7\ nm}{cm} \qquad \textit{Equation 4.6}$$

Peak Frequency

The vibrational modes and frequencies of a molecule are characteristic of its constituent nuclei and bond strengths. It follows that the Raman scattering frequencies of a material, and hence the Raman spectra peak positions, are characteristic of its composition. Above ~1800 cm^{-1} the assignment of Raman bands to particular molecular vibrational modes is relatively straightforward, since these modes tend to be localised to particular bonds or functional groups. Vibrational modes giving rise to Raman bands below ~1800 cm^{-1} are much more complex, involving more of the constituent nuclei and thus being more characteristic of the whole molecule rather than individual groups within it. The region from ~500 to ~1800 cm^{-1} is thus commonly known as the 'fingerprint region' [94]. The Raman shifts of some common vibrational modes are shown in Figure 4-5.

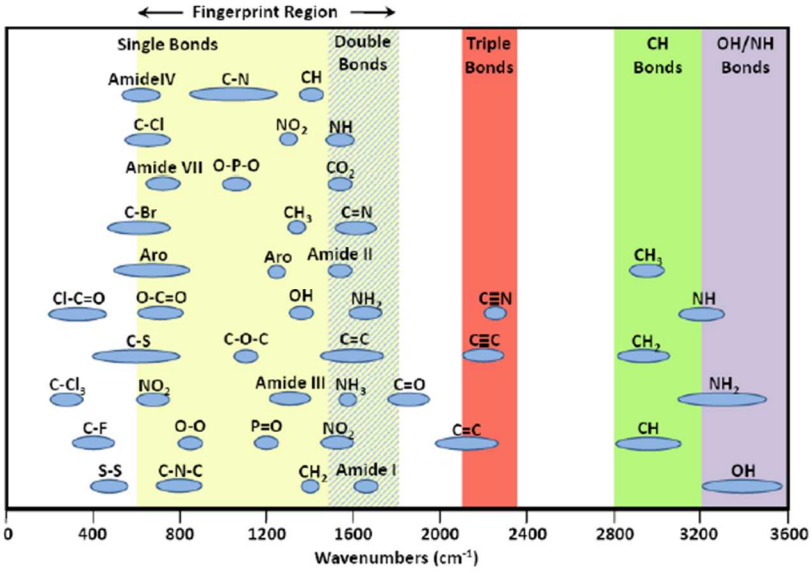


Figure 4-5 Raman shifts of some common bond vibrations. Reproduced from [95].

Peak Intensity

The intensity of Raman scattered light caused by a laser incident on a sample is proportional to the laser intensity, fourth power of the laser frequency, the number of scattering molecules in the sample, and the square of the polarisability, as shown in Equation 4.7 [90],

$$I_R \propto \nu^4 I_0 N \left(\frac{d\alpha}{dQ}\right)^2 \qquad \textit{Equation 4.7}$$

where I_R is the intensity of Raman scattered radiation, I_0 and ν are the incident laser intensity and frequency respectively, N is the number of scattering molecules, α is the polarisability of the molecules and Q is the vibrational amplitude. Therefore, the key theoretical factors affecting the intensity of peaks in a Raman spectrum include the amount of sample irradiated, and the particular laser (power and frequency) used as the excitation source.

A Raman spectrometer fundamentally measures intensity by counting the relative number of photons striking a charge-coupled device (CCD) following 'collection' of the Raman-scattered light via the instrument optics. Therefore, the 'collection efficiency' of the optics to channel Raman scattering to the detector is also important, and this in turn can be affected by the alignment of the instrument, and the optical properties and presentation of the sample being investigated. Furthermore, the 'quantum efficiency' of a CCD describes its sensitivity to light i.e. how many photons need to strike the detector to produce a measurable electrical

response. It should be noted that quantum efficiency varies with wavelength and will also inevitably vary between instruments.

Finally, if a particular sample fluoresces, this can produce a signal that overlaps and dominates the Raman spectrum. This is due to the fact that fluorescence is an absorption-emission process that is typically more intense than Raman scattering. Fluorescence occurs when the incident radiation is close in energy to the difference between two electronic states in the sample. Biological samples often possess some intrinsic fluorescence and in these cases employing a laser with relatively low energy, low frequency (near infrared or infrared) light can reduce fluorescence emission. This has the added benefit of reducing the likelihood of heat damage to the sample. According to Equation 4.7, however, the intensity of Raman-scattered light varies with the fourth power of the incident light frequency. With low frequency infrared light the Raman effect is thus much weaker with respect to that arising from illumination with visible or ultraviolet light. The selection of laser frequency to use for a particular sample is therefore a trade-off between wanting to maximise Raman scattering, but minimise potential heat damage or fluorescence effects. In this work a near-infrared (785nm) laser was used.

Peak Shape

Relatively few studies employing Raman spectroscopy consider peak shape in addition to peak position and intensity. The shape of a Raman band can, however, can yield information about the local environment of a vibrational mode [96]. For example, Izumi and Temperini found that the full width at half maximum (FWHM) of a band at ~1725 cm^{-1} in the spectra of polyhydroxyalkanoates was sensitive to the crystallinity of the sample [97]. Polyhydroxybutyrate, the more crystalline sample, gave rise to a band with a narrower FWHM than that arising from the more amorphous copolymer polyhydroxybutyrate-co-hydroxyvalerate.

4.1.4 Confocal Raman Micro-Spectroscopy (Raman Microscopy)

In confocal Raman microscopy, a Raman spectrometer is coupled with a confocal microscope to allow very small sample volumes (in the order of microns) to be targeted. The Renishaw instrument used in this work is pseudo-confocal in that instead of a conventional pinhole, it uses a slit and binning of the CCD pixels (i.e. combining several pixels into one pixel) to reject out-of-plane light. The slit is a one-dimensional spatial filter that allows only

light scattered from the focal point of the laser spot to pass through. In combination with this, the instrument computer is programmed to process and bin together data only from selected CCD pixels in the vertical dimension. The combination of the slit with the CCD pixel selection and binning thus creates a full confocal-like effect. The main advantage of this arrangement is that it is much easier to align a slit than a pinhole. More detailed descriptions of this pseudo-confocal arrangement can be found in the literature [98, 99].

4.2 Acquisition of Raman Spectra

4.2.1 Instrumentation

Raman spectra were acquired with an InVia Raman microscope (Renishaw plc, UK) coupled to an upright Leica DM LM microscope (Leica Microsystems, Germany) equipped with a HCX PL Fluotar 100×/0.9 NA objective (11566202, Leica Microsystems, Germany). Images of this instrument, together with a schematic diagram of the key components and optical paths, are presented in Figure 4-6.

Samples were irradiated with a 400 mW, 785 nm diode laser (Toptica, Germany) after positioning using a precision-motorised XYZ stage (Prior Scientific, UK). The power of the laser measured at the sample was 50 mW – much less than the rated 400 mW. This is attributed firstly to the age of the laser and associated power degradation. The second reason for this is the cumulative attenuation of power by each optical component in the laser path. Laser power measured at the sample was approximately 50% less than that measured before the beam expander.

Raman-scattered light from each sample was separated into its constituent wavelengths i.e. dispersed by a 1200 lines/mm diffraction grating. This light was projected onto a Peltier-cooled, 578 x 385 pixel, RenCam CCD (charge-coupled device) that measured the intensity of light at each wavelength. The CCD readout was processed by the instrument-provided software, WiRE 2.0 (Renishaw plc, UK), to yield a Raman spectrum.

The Leica microscope was equipped with a 100 W halogen lamp for widefield illumination of samples. It was found, however, that this lamp caused significant localised heating. As a result, stable positioning of a small sample such as a bacterium proved problematic due to thermal expansion and contraction: Whilst the lamp was switched on to identify and position a sample, localised heating and thermal expansion would cause systematic movement of the

sample relative to the objective (Figure 4-7). When the lamp was subsequently turned off to acquire a Raman spectrum, localised cooling and contraction would move the selected sample away from alignment with the laser path. For this reason a cold light source – an LED array lamp – was used to illuminate and position samples for subsequent Raman analysis.

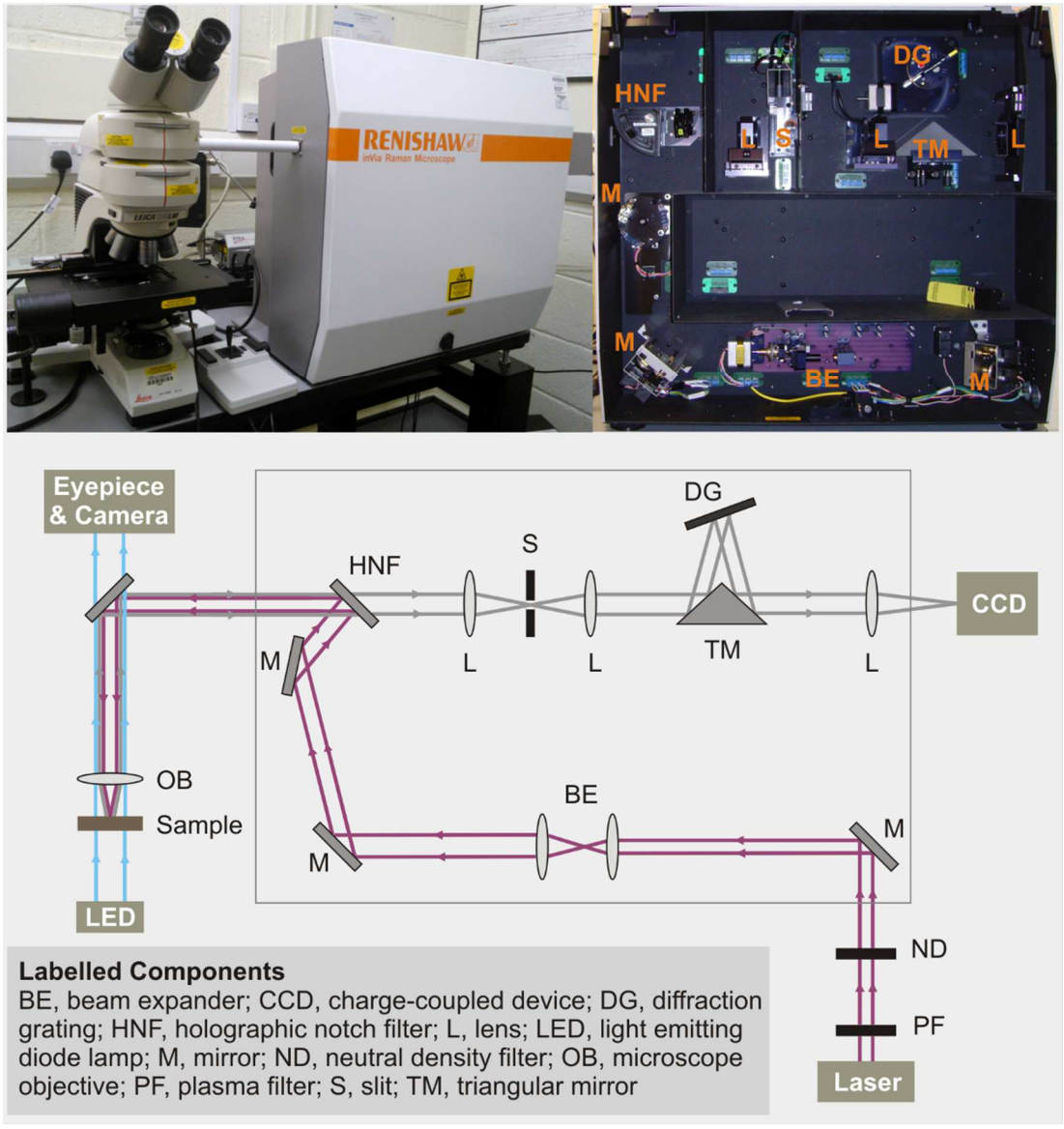


Figure 4-6 The Renishaw InVia Raman microscope used in this work (top left); View of the components inside the Raman instrument (top right); and schematic diagram illustrating key components and optical paths through the instrument (bottom).

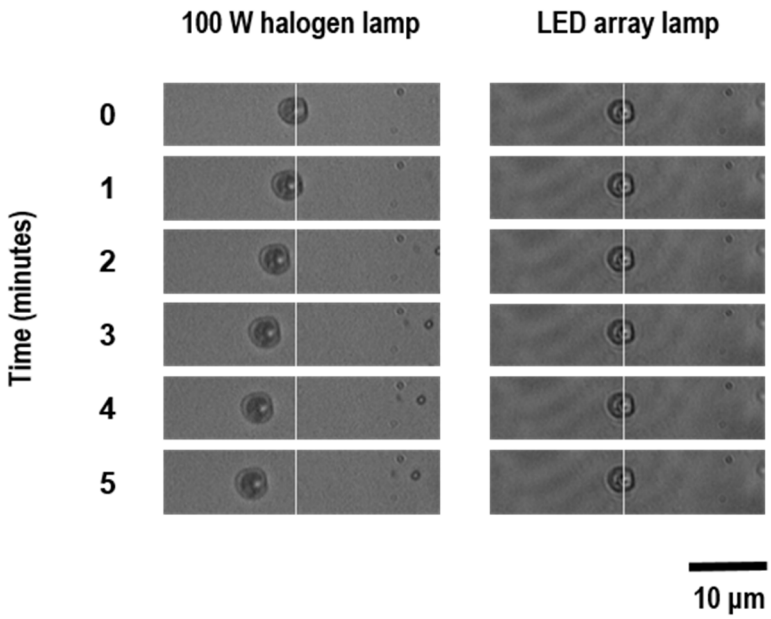


Figure 4-7 Change in position of a small object as a function of time when illuminated by a hot light source (100 W halogen lamp) versus a cold light source (LED array lamp).

4.2.2 Spectral Resolution

Spectral resolution is a measure of the degree to which two closely positioned peaks may be distinguished from one another. It is expressed in either a frequency or wavenumber that represents how close two peaks may be to each other and still be identified as distinct peaks rather than combined as a single peak. It should be noted that for a dispersive Raman instrument (i.e. one that uses a dispersion grating and a multi-channel detector) spectral resolution is uniform across all wavelengths but varies with Raman shift. This is due to the fact that the relationship between the wavelength of scattered light and Raman shift in wavenumbers is not linear (see Section 4.1.3).

For the set-up of the Renishaw InVia Raman microscope described here (785 nm diode laser and 1200 lines/mm dispersion grating), the step size (in wavenumbers) of collected data points was 0.9383 cm^{-1}. To gain an approximation for spectral resolution a spectrum of a neon light source (Renishaw plc, UK), was taken since this has multiple narrow spectral lines distributed throughout the typical spectral range. For several of these bands, the full width at half maximum (FWHM) was determined and plotted versus spectral position (Figure 4-8). For the range of wavenumbers routinely investigated in this work, it can be seen that the FWHM (and hence we approximate that the spectral resolution) varied from 3.2

cm^{-1} (at 400 cm^{-1}) to 2.6 cm^{-1} (at 1800 cm^{-1}) with a standard deviation of 0.07 cm^{-1} determined from the linear regression statistics.

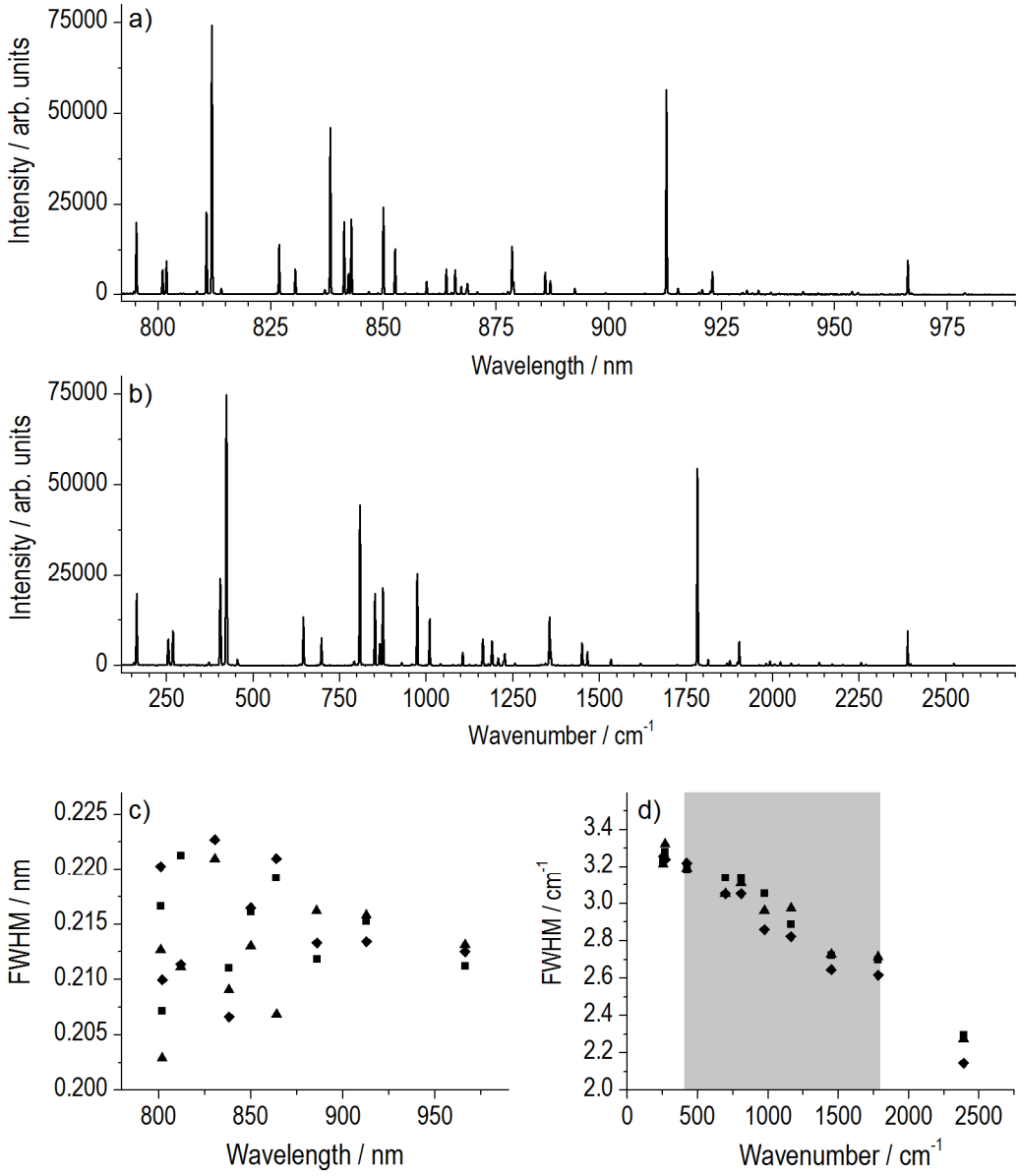


Figure 4-8 Raman spectra of neon for experimental determination of spectral resolution. The abscissa is displayed in wavelength of scattered light (a) and "Raman shift" in wavenumbers (b). Full width at half maximum (FWHM) for several neon spectral bands are plotted against wavelength (c) and wavenumber (d) to illustrate how spectral resolution has a frequency dependence when reported in wavenumbers. The different symbols in (c) and (d) represent FWHM values determined from three separate neon spectra. The shaded grey area in (d) shows the range of wavenumbers routinely considered in this work.

Typically, a measure of spectral resolution made this way will overstate the true spectral resolution. This is because neon gives a spectrum of atomic emission lines rather than a true Raman spectrum of inelastically scattered light. It does not, therefore, take into account the effect of the laser bandwidth imposed on spectral resolution [100]. For the purpose of this work, however, this approximation of the spectral resolution is sufficient.

4.2.3 Spatial Resolution

Spatial resolution in Raman spectroscopy describes the minimum distance required between two points for them to be detected as separate entities. The theoretical values for lateral (in-plane) and axial spatial resolution are given below for the instrumental set-up used in this work.

Lateral Spatial Resolution

Theoretical diffraction-limited lateral spatial resolution for a confocal microscope is given by the expression:

$$\frac{0.61\,\lambda}{NA} \qquad \textit{Equation 4.8}$$

where NA is numerical aperture of the lens and λ is the wavelength of incident light [101]. For the system described here, λ = 785 nm and NA = 0.9. The theoretical lateral spatial resolution is therefore 0.5 µm. With 'good' samples – where other optical process such as scattering and refraction are negligible, spatial resolution approaching the diffraction limit may be achieved. For samples not satisfying this criteria, however, spatial resolution is typically quoted to be approximately 1 µm [102].

Axial Spatial Resolution

A confocal microscope uses a pinhole to reject out-of-focus light, effectively yielding a much smaller axial sampling volume (hence a much greater axial spatial resolution) than can be achieved with a conventional microscope. The Renishaw InVia Raman microscope used in this work is a pseudo-confocal instrument, using a slit rather than a pinhole to achieve confocality.

Theoretical axial spatial resolution is given by the expression:

$$\frac{2n\lambda}{(NA)^2} \qquad \textit{Equation 4.9}$$

where n is the refractive index of the sample, NA is numerical aperture of the lens and λ is the wavelength of incident light [103]. The refractive index of air is 1.00, and for the system described here, λ = 785 nm and NA = 0.9. The theoretical axial spatial resolution is therefore 1.9 µm.

4.2.4 Experimental Estimation of Raman Sampling Volume

Experimental estimation of Raman sampling volume was made from Raman line maps of various materials (specified below) with static scans being acquired at 0.2 µm step intervals for in-plane (lateral) scans and 1 µm step intervals for axial scans. Data collected from Raman line scans were analysed as follows: The peak of interest for each material (at approximately 320, 520, 877, 1000 and 1654 cm^{-1} for CaF_2, silicon, aqueous NaH_2PO_4, polystyrene and rapeseed oil respectively) was fit with a smoothing spline and integrated in MATLAB Release 2014a (The MathWorks Inc., MA, US). Integrated peak areas were plotted as a function of lateral displacement to yield a 'response curve' representing the change in band intensity as the material interface moved across the path of the laser. The first derivative of these response curves was calculated (also in MATLAB) and subsequently fit using both 'GaussAmp' (Equation 4.10) and 'InvsPoly' (Equation 4.11) functions in Origin software, where y_0 = offset, x_c = centre, w = width, A = amplitude, and A_1, A_2 and A_3 are all coefficients. For both fit functions, user-defined constraints for y_0 (zero) and A (peak height determined by visual inspection) were employed. A diagrammatic representation of curve-fit parameters is shown in Figure 4-9. The reported FWHM from these curve-fits was used directly as an approximation for laser spot size and hence Raman sampling volume.

$$y = y_0 + Ae^{-\frac{(x-x_c)^2}{2w^2}} \qquad \textit{Equation 4.10}$$

$$y = y_0 + \frac{A}{1+A_1(2\frac{x-x_c}{w})^2+A_2(2\frac{x-x_c}{w})^2+A_1(2\frac{x-x_c}{w})^6} \qquad \textit{Equation 4.11}$$

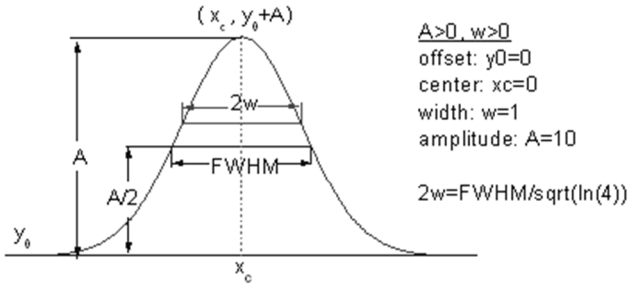


Figure 4-9 Sample curve for 'GaussAmp' curve fitting function in Origin, diagrammatically showing parameters relevant to both 'GaussAmp' and 'InvsPoly' fits. Taken from [104].

Determination of the in-Plane (Lateral) Focussed Laser Spot Size

The results for one experiment determining the spot size of the Raman laser are shown in Figure 4-10. In this case, a 5 µm polystyrene sphere deposited on a quartz substrate was passed through the path of the laser in 0.2 µm steps. The intensity of the Raman spectrum changes in response to the relative position of the polystyrene sphere and the focussed laser spot and this is measured using the strong band at 1000 cm^{-1} (Figure 4-10b). As the sphere enters the laser spot, the Raman spectral intensity increases up to a maximum that corresponds with the centre of the sphere. With continuation of the line map, the intensity then decreases to zero as the sphere passes out of the laser spot (Figure 4-10c). By visual inspection, the FWHM of the response curve is approximately 6 µm – slightly wider than the stated diameter of the polystyrene sphere (5 µm). The derivative of the response curve has two peaks, each representing one 'edge' (i.e. interface) of the sphere (Figure 4-10d). The fits of the GaussAmp and InvsPoly functions to these peaks are similar in the region of the FWHM. At the base of the curves at the extreme points, however, by visual inspection GaussAmp fits this particular data set slightly better than InvsPoly. The correlation coefficients for the GaussAmp and InvsPoly fits versus the experimental data are 0.946 and 0.957 respectively, indicating that both functions provide a good fit and that there is little difference between them. For each spot size measurement, the reported FWHM from both

peak fits (representing the two interfaces) were averaged to give one FWHM for each curve fit function.

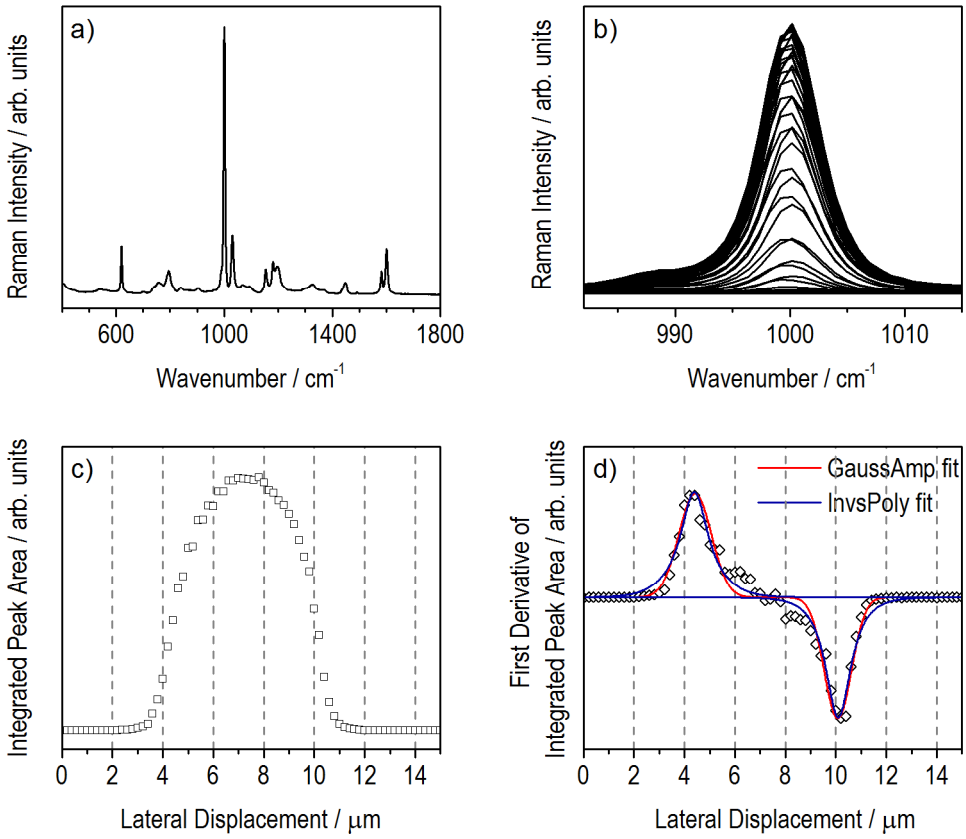


Figure 4-10 Experimental estimation of laser spot size in the lateral plane using 5 μm polystyrene spheres deposited on a quartz substrate. A Raman line map through a single polystyrene sphere was performed, with spectra acquired as a function of displacement along this line. The strongest peak at 1000 cm^{-1} (a) was used for subsequent analysis. Peak intensities of all the acquired spectra (b) were integrated and plotted as a function of lateral displacement (c) to give a 'response curve'. This profile represents the two edges of the sphere as it moves through the path of the laser. The first derivative of the response curve was calculated (d) and the FWHM of both peaks (each representing one edge of the sphere) was averaged to yield an approximation of the laser spot size.

The spot size measurement was carried out three times along the x-axis and three times along the y-axis of the stage motor. The FWHMs were averaged and are shown in Table 4-1. The FWHM from x-axis and y-axis GaussAmp fits can be averaged to yield a FWHM, and hence a lateral spot size of 1.2 µm.

Table 4-1 Experimental parameters and FWHM determined from Raman line maps through 5 µm polystyrene beads in the lateral plane. The reported data statistics were calculated from three measurements made along both the x-axis and y-axis.

Direction	Approx. Peak Centre / cm^{-1}	Acquisition Time * / sec	FWHM / µm (GaussAmp fit)			FWHM / µm (InvsPoly fit)		
			Mean	Std. Dev.	R^2	Mean	Std. Dev.	R^2
x-axis	1000	3	**1.1**	0.1	0.9	**1.0**	0.1	0.9
y-axis	1000	3	**1.3**	0.3	0.9	**1.2**	0.2	1.0
* per individual spectrum								

Determination of the Axial Range of the Focussed Laser Spot

The axial range of the focussed laser spot was estimated from Raman depth scans through several different material interfaces. The scans were started at depth within the material and moved toward the interface with the minimum possible step size of 1 µm. The results for one measurement per material – including Raman spectrum, response curve (integrated peak area as a function of axial displacement), and first derivative of response curve – are shown in Figure 4-11.

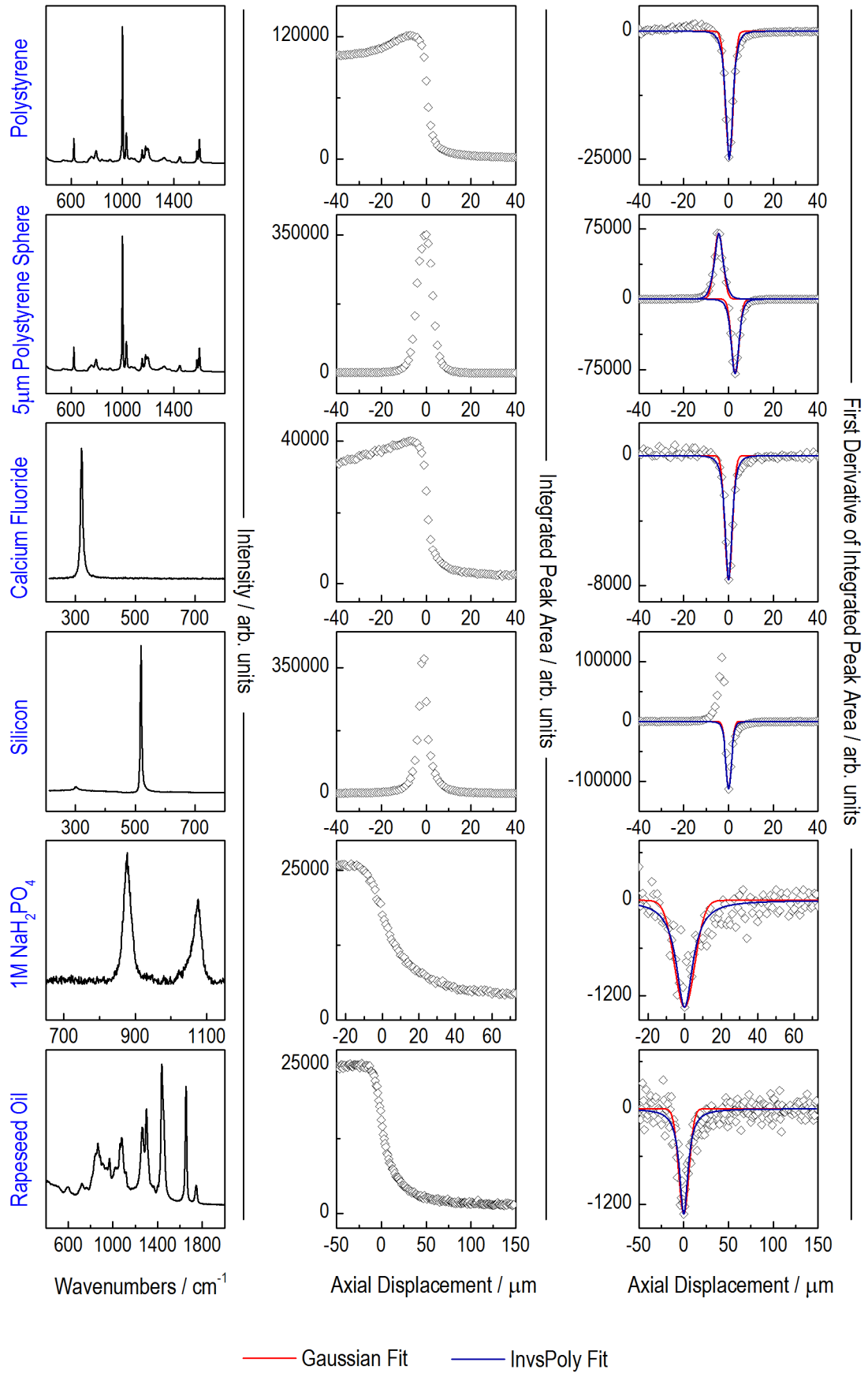

Polystyrene
5µm Polystyrene Sphere
Calcium Fluoride
Silicon
1M NaH_2PO_4
Rapeseed Oil
Intensity / arb. units
Wavenumbers / cm^{-1}
Integrated Peak Area / arb. units
Axial Displacement / µm
First Derivative of Integrated Peak Area / arb. units
Axial Displacement / µm
Gaussian Fit
InvsPoly Fit

Figure 4-11 Experimental estimation of the axial range of the focussed laser spot using different materials. Raman depth scans through each material were performed, with spectra acquired as a function of displacement from the material interface. A strong peak from each Raman spectrum (left) was used for subsequent analysis. Peak areas were integrated and plotted as a function of axial displacement (middle) to give a 'response curve'. The first derivative of the response curve was calculated (right) and the FWHM determined from fitting with a Gaussian or InvsPoly function to give an approximation of the axial laser spot range.

Three shapes of response curve are immediately apparent; first, a slight increase in intensity toward a maximum followed by a sharp drop in intensity to zero (observed for polystyrene and CaF_2). Second, a constant non-zero intensity followed by a drop in intensity to zero (observed for 1M NaH_2PO_4 and rapeseed oil); and third, a full peak shape (observed for the 5 μm polystyrene sphere and silicon). A peak shape is expected in the response curve for the polystyrene sphere since the focussed laser spot encounters two interfaces as the sphere tracks through it. For silicon, however, the peak shape in the response curve is due to the fact that the material strongly absorbs the radiation so that the penetration depth of 785 nm light is just 12 μm [105]. For polystyrene and CaF_2, the slight drop off in signal intensity as the laser focal spot goes deeper into the material could be due to the difference in refractive index of air (~1) and the material (e.g. 1.4 for CaF_2). By contrast, the refractive indices for the aqueous and liquid samples (e.g. 1.33 for water) are much more closely matched to the refractive index of the magnesium fluoride (MgF_2) encountered at the interface.

It can be seen that the fits of the GaussAmp and InvsPoly functions to the derivative of the response curves are similar in the region of the FWHM. At the base of the curves, however, by visual inspection InvsPoly fits these data slightly better than GaussAmp. The spot size measurement was carried out three times per material. The FWHMs were averaged and are shown in Table 4-2. Two groups of results are apparent: those for the solid-air interface measurements (polystyrene, polystyrene sphere, CaF_2, Silicon) and those for aqueous or liquid interface with a solid MgF_2 window (1M NaH_2PO_4, rapeseed oil). The results of the former can be averaged to give an FWHM and hence axial resolution of 4 μm. Note that the result from silicon was not included in this calculation due to the strong absorption effect of the material. The results for 1M NaH_2PO_4, and rapeseed oil can be averaged to yield an axial spatial resolution of 11 μm. Note that the greater standard deviations in this result are likely due to the lower signal to noise ratio of the Raman spectra obtained for these materials and subsequent poorer curve-fits of the data.

Table 4-2 Experimental parameters and FWHM determined from Raman line scans through different materials in the axial plane. The reported data statistics were calculated from three measurements made on each material.

Sample	Approx. Peak Centre / cm^{-1}	Acquisition Time * / sec	FWHM / μm (Gaussian fit)			FWHM / μm (InvsPoly fit)		
			Mean	Std. Dev.	R^2	Mean	Std. Dev.	R^2
Polystyrene	1000	1	**4.1**	0.3	0.91	**3.7**	0.3	0.93
5μm polystyrene bead on quartz	1000	3	**4.1**	0.4	0.99	**4.1**	0.4	0.99
Calcium Fluoride	320	1	**4.0**	0.2	0.95	**3.7**	0.1	0.97
Silicon	520	0.2	**3.1**	0.1	0.95	**3.1**	0.2	0.91
1M NaH_2PO_4 **	877	10	**11.1**	3.2	0.48	**9.2**	2.9	0.57
Rapeseed oil **	1654	3	**12.7**	0.7	0.67	**10.6**	0.8	0.72

* per individual spectrum

** enclosed in a small well with a MgF_2 window

A large contributor to the difference between spot range measured for the solid materials (4 μm) and the aqueous or liquid materials (11 μm) is likely to be the difference in the optical paths travelled by the laser for the two types of sample. This means that the laser is subject to different refractive, absorptive and scattering effects. In the former type of sample, the laser encounters the surface of the material directly, but is subject to a large difference in refractive index at the interface of the depth-scan measurement. In the latter type of sample, the laser first passes through an MgF_2 window, through which greater absorptive and scattering effects could occur, before encountering the interface over which the depth scan is conducted. The refractive index of the materials on either side of this interface are more closely matched to each other than to air.

Estimation of Raman Sampling Volume

The experimentally-determined laser spot size was 1.2 μm in the lateral plane. The axial range of the spot was determined to be 4 μm for the material-air interfaces, and 11 μm for

the solution or liquid/MgF_2 interfaces investigated. Ideally these measurements would have been made on samples that are smaller than the laser spot size – i.e. very small microspheres or very thin samples. The samples used here (due to availability) were larger than the laser spot size and therefore, the measured Raman intensity will have some contribution from material excited by out-of-focus regions of the laser spot. The laser spot size is thus likely to be overestimated by this method. Furthermore, the samples used here will have rough surfaces with respect to an optically polished surface. This can lead to scattering of the incident laser and poor collection of the Raman-scattered radiation with respect to an optically smooth surface. This, however, reflects the nature of the samples being measured with this method and so is reasonable. Finally, it should be noted that for our instrument, the step motor resolution in the axial plane is low (1 μm) relative to the resolution in the lateral plane (0.1 μm). The stated accuracy in the estimated spot size reflects this.

Raman sampling volume can be approximated by laser spot volume [106]. For simplicity, this is assumed to be a cylinder. Using our experimentally determined dimensions, the sampling volume is therefore estimated to be 4.5 μm^3 for the material-air interfaces and 12.5 μm^3 for the solution/liquid-MgF_2 interfaces.

4.2.5 Instrument Alignment

The alignment of the Raman system was checked each day before acquiring any Raman spectra. First, a silicon sample (Renishaw plc, UK) was used to visualise the laser spot position in the camera viewing window and re-centre it if necessary. This silicon sample is mounted in a defined crystallographic direction so this did not change between measurements. With the laser spot focused on the surface of the silicon, the WiRE software was then used to auto-align both the CCD camera and the instrument slits. Finally, the 'Quick Cal' feature in WiRE was used to align the wavenumber axis. In 64 silicon spectra recorded after calibration on different days between January 2014 and September 2015, the mean position of the major peak was found to be at 520.2 cm^{-1} with a standard deviation of 0.1 cm^{-1}.

4.2.6 Acquisition Parameters

Raman spectra were acquired between 400 cm^{-1} and 1800 cm^{-1} for each sample. This region was selected since it encompassed the Raman fingerprint region as well as the most extreme peaks of interest for polyhydroxyalkanoates (430 and 1725 cm^{-1}). A 'static' Raman spectrum acquired on this system with a 1200 l/mm grating has a spectral scan width of approximately 550 cm^{-1}. The SynchroScan™ method unique to the Renishaw instrument was thus used to obtain a continuous 'extended scan' that covered the required range of wavenumbers. Note that if an extended scan time of 10 seconds was input into the instrument, the scan took 37 seconds to acquire in real time. This was scalable so that a 200 second extended scan took approximately 12 minutes to acquire in real time. A number of different acquisition times were used in this work so they are reported individually for each data set.

Raman point-spectra rather than maps were acquired from each sample cell. A typical bacterium was 2 to 5 μm in size, whereas the minimum achievable lateral spatial resolution of the instrument was calculated above to be 0.5 μm. The minimum achievable laser spot size is equal to twice the lateral spatial resolution (i.e. 1 μm), and even then, regions of the sample that are outwith the laser focus would be excited to some degree. The relative sizes of the bacteria and the laser spot meant that for most cells, it is likely that the acquired Raman spectra represented a subsample of the whole cell. Figure 4-12 shows the Raman spectra from line maps acquired through two different samples. It can be seen that the strongest signal in the top example is obtained from the centre of the bacterium. It can also be seen that although there is variation in spectral intensity the components in the spectrum do not appear to change. If the size of the laser spot was similar to the size of the cell, the drop-off in spectral intensity could be due solely to the laser spot increasingly encompassing regions of the substrate with no cellular material. In the bottom example, the strongest signal again, is obtained from the centre of the cells and almost no signal is recorded in the centre of the scan at the point where the two cells are dividing. Bearing in mind this likelihood of subsampling, to achieve some degree of consistency, Raman spectra were purposefully acquired from as close as possible to the centre of each cell measured in this work.

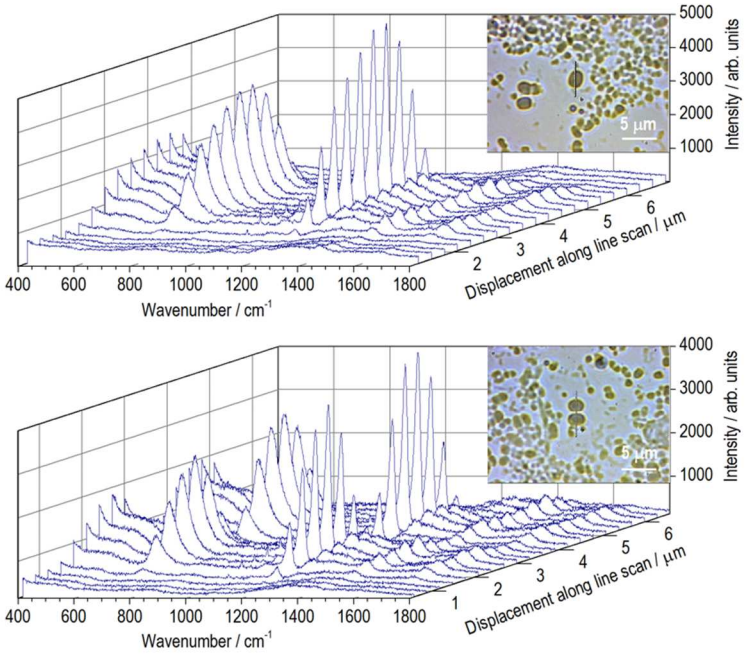


Figure 4-12 Raman line maps through two example bacterial cells: a coccus-shaped organism (top) and a coccus-shaped organism that appears to be dividing (bottom).

4.2.7 Sample Preparation

Bacterial samples were prepared according to the method by Majed *et al*. [68]. In brief, 1 mL samples of EBPR mixed liquor (bacterial suspension) were collected and aspirated with a syringe and 26 gauge needle to disrupt bacterial flocs. Small drops of each sample were deposited on CaF_2 substrates and left to dry in air before subsequent Raman analysis.

Aqueous polyphosphate samples were pipetted into custom-made aluminium microscope slides (75 × 30 × 2 mm) containing one small (2 mm diameter × 1 mm depth) well each. Each well was sealed with a glass coverslip that was secured in place using nail varnish. The Raman laser was focused 50 µm into the sample from the bottom of the coverslip to minimise the contribution of the glass coverslip to the Raman spectrum (Figure 4-13).

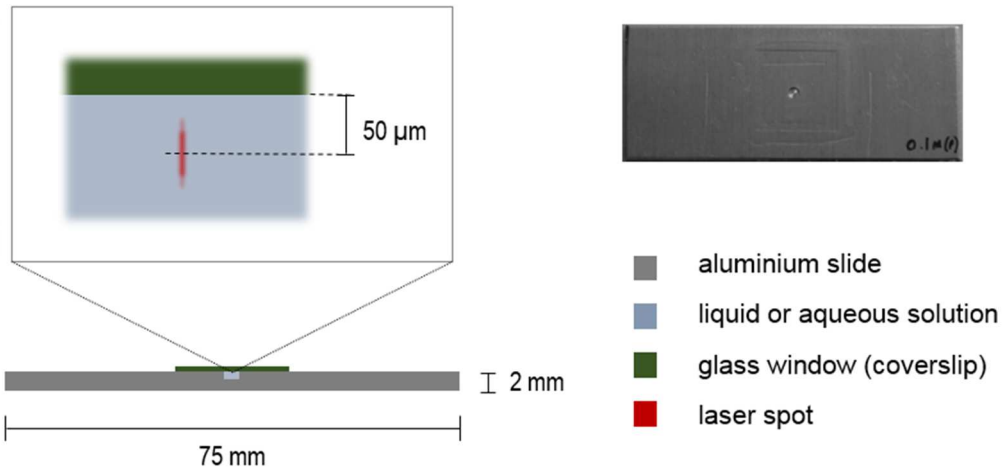


Figure 4-13 Schematic diagram (left, not-to-scale) and image (top-right) of a custom made aluminium slide for presentation of aqueous samples. See text for details.

4.3 SPECTRAL PREPROCESSING

4.3.1 Cosmic Spike Removal

Cosmic rays are ubiquitous, high energy particles that can randomly strike the CCD during acquisition of a Raman spectrum. They cause high intensity, narrow bandwidth signals in a Raman spectrum which are referred to as cosmic spikes. Since these spikes overwhelm a spectrum, it is necessary to remove them before further analysis. Although some automated algorithms exist to remove cosmic spikes, these were found to be unsuitable for the type of spectra investigated here. Cosmic spikes were therefore removed from each Raman spectra individually using the zap function in the WiRE software.

4.3.2 Background Removal and Baselining

In addition to the signals of interest, the Raman spectra measured had contributions from the substrate on which the samples were prepared (in most cases CaF_2), and also autofluorescence arising from the bacterial samples. The intensities of these contributions vary uncontrollably and yield no useful physicochemical information about each sample. It was necessary, therefore, to remove these contributions during preprocessing so the spectra from different samples could be compared directly. Using MATLAB and R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria) [107], the best way to achieve this was investigated and is the subject of Chapter 5.

4.4 Overview of Spectral Analysis

4.4.1 General Methods

A number of different processing methods were used so they are reported individually in each chapter. General data processing, to perform routines such as extracting peak height data or calculating derivatives, was performed in MATLAB. Curve fitting was carried out using Origin (OriginLab, Northampton, MA).

4.4.2 Multivariate Methods

Four multivariate statistical methods, namely principal components analysis (PCA), multiplicative curve resolution (MCR), partial least squares (PLS) regression, and partial least squares-discriminant analysis (PLS-DA), were used to analyse Raman spectra in various sections of this work.

PCA reduces the dimensionality of a data set by constructing a new set of orthogonal variables, or 'principal components' (PCs), that capture covariance in the original data set variables. By applying PCA, the major trends in the data may be described using far fewer variables than present in the original data. For example, the Raman spectra acquired in this work typically comprised ~1490 data points i.e. variables. Performing PCA on sets of spectra meant that usually, > 95% variance between spectra could be described using < 5 PCs. The PCs are ordered according to how much variance (i.e. information) in the data they capture i.e. the first PC captures the most variance and the remaining PCs follow in descending order. The analysis yields three key parameters: 'eigenvalues' – which describe the variance captured by each PC, 'loadings' (or 'loads') – which are a linear combination of the original variables present in each PC, and 'scores' – which for each sample in the data set describes how much of each PC is present.

MCR, also known as end member extraction, aims to extract underlying pure component spectra and related concentrations from a data set with as few assumptions about the data as possible [108, 109]. Unlike PCA, the extracted components are not constrained by a requirement to be orthogonal. They do, however, have a non-negative constraint which aids the algorithm in reaching a solution and in providing components that are (hopefully) physically meaningful.

Partial Least Squares (PLS) regression is a type of multivariate regression. In general, a regression model seeks to relate a set of dependent ('X-block') variables which are easily

measureable to a set of independent ('Y-block') variables which constitute a property of interest such as concentration which is not itself directly measurable. The regression model can then be used to predict the property of interest in future samples. PLS regression is similar to PCA regression. In the latter case, principal components rather than the original variables are used as the X-block variables. As described above, however, principal components are determined independently of the Y-block data. This means that although PCs capture the major trends in the X-block data, these may not be the most useful factors for predicting the properties of interest. By contrast, PLS regression builds a model that maximises variance captured in X that is correlated to (i.e. useful for predicting) Y. The new set of (X) variables constructed in PLS models are known as latent variables (LVs) rather than principal components.

PLS-DA is a particular type of PLS regression, where the Y-block data comprises class-type (discrete) data rather than continuous variables.

All multivariate analyses were carried out with PLS_Toolbox (Version 7.0.3) for use with MATLAB (Eigenvector Research, Inc, Manson, WA).

Chapter 5
Baselining Raman Spectra of Individual EBPR Bacteria

5.1 INTRODUCTION 87
5.1.1 Character of EBPR Bacteria Spectra 88
5.1.2 Approaches to Baselining 90
5.1.3 Selection of Methods for Comparison 94
5.2 METHODOLOGY 95
5.2.1 Spectroscopic Measurements 95
5.2.2 Data Sets 95
5.2.3 Preprocessing Methods 95
5.2.4 Evaluation of Baselining Methods 103
5.3 RESULTS AND DISCUSSION 104
5.3.1 Appearance of Corrected Spectra 104
5.3.2 Potential Effect on Classification 109
5.3.3 Potential Effect on Quantification of Polyphosphate 116
5.3.4 Other Considerations 118
5.3.5 Summary and Conclusions 119

5.1 INTRODUCTION

Baselining is one kind of preprocessing which in most cases involves ‘flattening’ a spectrum so that all points representing zero sample signal actually lie along the line of zero intensity rather than being positively shifted. In some cases the baseline contribution is normalised rather than removed. The purpose of baselining Raman spectra is to remove variation that does not reflect true chemical or physical changes in the samples themselves. It is vital for robust analysis and accurate interpretation of the available data.

Uncontrollable variation between spectra may arise from several different sources. Fluorescence in particular, is recognised as a significant contributor to Raman spectra, giving rise to broad background signals. In addition, dark counts are a type of instrumental contribution that arise from the accumulation of charge by a CCD detector over time even when there is no light striking it. Finally, for some samples, the substrate on which they are presented or the matrix in which they are contained may give variable contributions to the acquired Raman spectra. Removal of these kinds of unwanted, uncontrollable variation between samples is necessary to enable meaningful analysis and interpretation of the spectral data.

It should be noted that in this study the terms ‘background subtraction’ and ‘baselining’ are used to indicate two distinct processes. Background subtraction is used to refer to the removal of a known background spectrum – such as that from a substrate or a sample matrix. Baselining is used to indicate the process of either normalising the baseline contribution or flattening a spectrum by mathematically constructing the baseline that should be subtracted to achieve this effect.

In this work, exploratory analysis of the Raman spectra of EBPR bacteria is undertaken in order to investigate potential application of this tool to EBPR research. Baselining the full spectra of EBPR bacteria is thus required rather than focussing on any particular band or set of bands. In addition, for processing large numbers of samples it is desirable that the baselining method employed is automated and that it can handle the variation present between spectra.

A large number of approaches to automated baselining of full spectra are described in literature. Few review or comparison studies exist, however, to aid navigation of the literature or decision-making over which method to employ for a specific case. The aims of this chapter are therefore threefold. First, a brief overview and framework for categorising

existing baseline correction methods is provided to aid navigation of the literature. Second, a number of baselining methods are applied to a set of Raman spectra of individual EBPR bacteria. This serves as a direct comparison study of the performance of a wide variety of automated baselining methods on real (as opposed to synthetic) Raman spectra. Finally, a decision is made regarding which method to use to baseline Raman spectra for the remainder of this study.

5.1.1 Character of EBPR Bacteria Spectra

Prior to a discussion on baselining techniques, it seems prudent to give a brief introduction to the character of EBPR bacteria Raman spectra. Four spectra of individual EBPR bacteria are shown in Figure 5-1, together with spectra of three key storage compounds (metabolites) identified in EBPR – polyphosphate, polyhydroxyalkanoate (PHA) and glycogen. Note that polyhydroxybutyrate (PHB) is a type of PHA.

Figure 5-1a shows that there are similarities between the spectra of individual EBPR bacteria, such as the spectral bands at ~1002, 1450 and 1660 cm^{-1}. There are also, however, some significant differences between the spectra. Principally, this is due to the fact that when any of the bacterial storage compounds associated with EBPR are present, they may dominate the spectra. This is demonstrated in Figure 5-1 where the Raman spectra of polyphosphate (b), PHB (c) and glycogen (d) are clearly represented in the bacteria spectra. Furthermore, extra peaks are sometimes present in the sample spectra (not shown). These may be due to other unknown contaminants present as a result of the EBPR process itself being subject to a lot of variation.

Figure 5-1f shows the instrument dark spectrum and a calcium fluoride (CaF_2) spectrum. The Raman-grade CaF_2 substrates used in this study are reported to have one sharp Raman spectrum peak at 321 cm^{-1} (Crystran Ltd., UK). Evidently, they also have a broad, low level signal that becomes apparent when samples – such as the individual bacteria analysed here – have low signal intensity. Since the relative contribution of the CaF_2 spectrum to an EBPR bacterium spectrum is uncontrollable, it is one component that needs to be removed during preprocessing.

Finally, it should be noted that the shape of the Raman spectral bands themselves vary from individual sharp peaks such as the phenylalanine peak at ~1002 cm^{-1}, to broad bands such as one of the polyphosphate peaks at ~698 cm^{-1}, to broad regions of overlapping peaks, such as between 1200 and 1500 cm^{-1}. This variation in peak shape may in itself may be a challenge

to some baselining techniques. Further discussion on particular peak assignments lie outside the scope of this chapter but are given in Chapter 6.

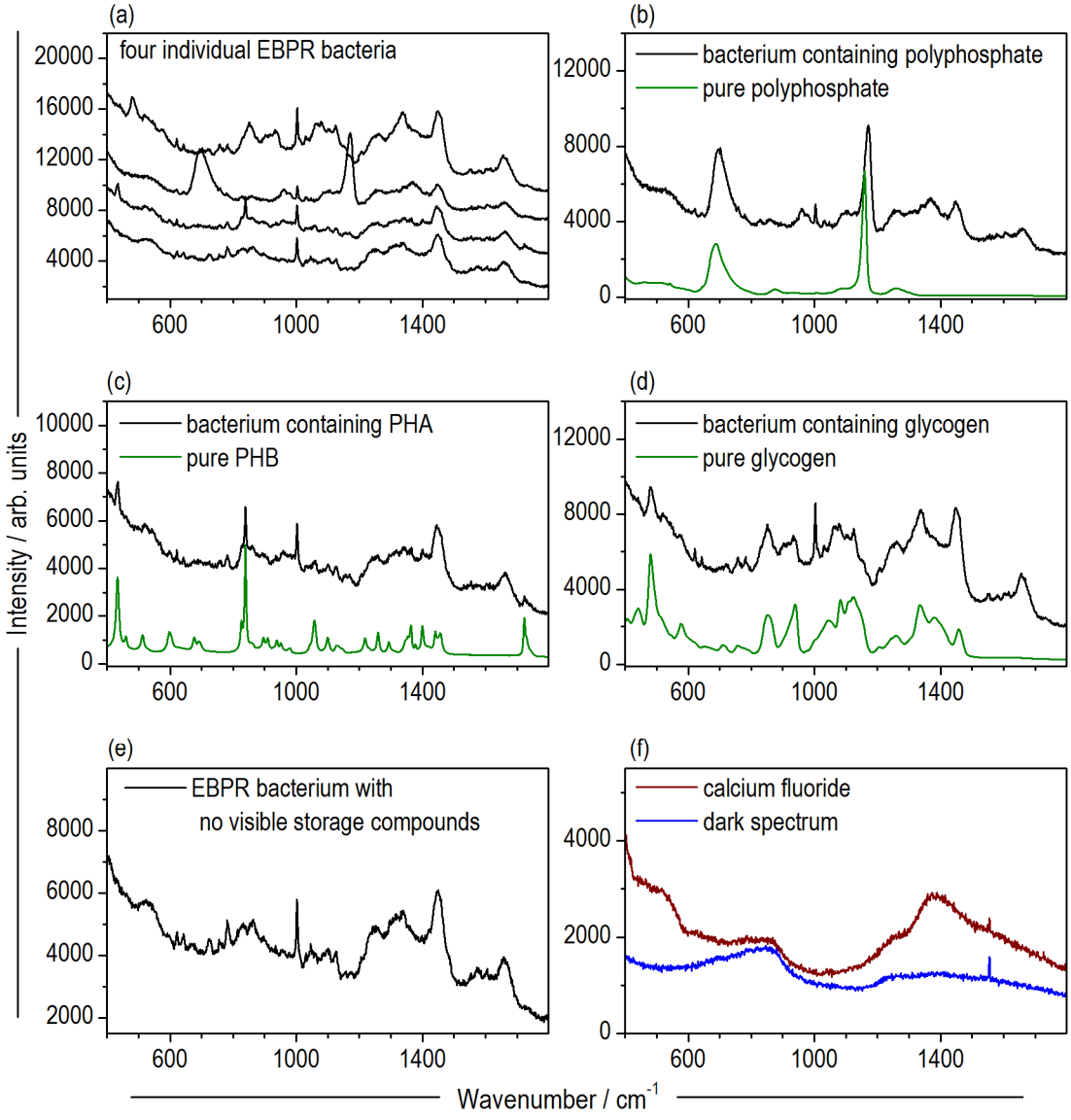


Figure 5-1 Raman spectra of four individual EBPR bacteria offset for clarity (a). The same spectra are shown individually together with a spectrum of the pure EBPR storage compound they contain – polyphosphate (b), PHB (a type of PHA) (c), glycogen (d), and no visible storage compound (e). The instrument dark spectrum and calcium fluoride substrate spectrum are shown in (f). Spectral acquisition time was 200 seconds for the bacteria, calcium fluoride and dark spectra, and 60 seconds for pure compounds. The intensities of pure compound spectra were scaled down for presentation.

5.1.2 Approaches to Baselining

A framework for baseline correction methods is presented in Figure 5-2. This is not intended to be a definitive classification but rather a starting point for categorising existing methods, with the aim of being a quick aid to navigating the literature. It can be seen this is not a perfect classification of methods in that some examples could belong to more than one category. The examples given in each category are not exhaustive and detailed descriptions of the methods are not provided except in the case that they are selected for comparison in this chapter. The reader may consult the provided references for further information.

Baseline correction methods may be first divided into those that normalise the baseline contribution, and those that actually remove the baseline. Normalisation techniques primarily standardise rather than remove the baseline, and include standard normal variate (SNV) [110], multiplicative scatter correction (MSC) [111] and extended multiplicative scatter correction (EMSC) [112]. These kind of techniques tend not to be suitable for cases where the spectral shape changes significantly from sample to sample.

Baseline removal techniques can be divided into those that treat each sample (i.e. spectrum) separately and those that model the baseline based on the information in a data set comprising several spectra. The latter include singular value decomposition (SVD) [113, 114], artificial neural networks (ANN) [115] and orthogonal signal correction (OSC) [116]. Applying a baselining model from one data set to a new data set would require that the full range of variation in the new data is represented in the original calibration model. With the large potential variation evident in EBPR bacteria Raman spectra, this is not an ideal scenario. Alternatively, if each data set were baselined by construction of a new model, it may be difficult to directly compare spectra from different data sets.

Three categories of ‘per sample’-based baselining are considered here: model-free methods, frequency filtering, and those that fit a function such as a spline or polynomial through selected data points. Model-free approaches include local medians (LM) [117] and derivatives [118, 119] (in the case that the original spectrum is not reconstructed from modeling the derivative). These methods are relatively simple and fast to calculate.

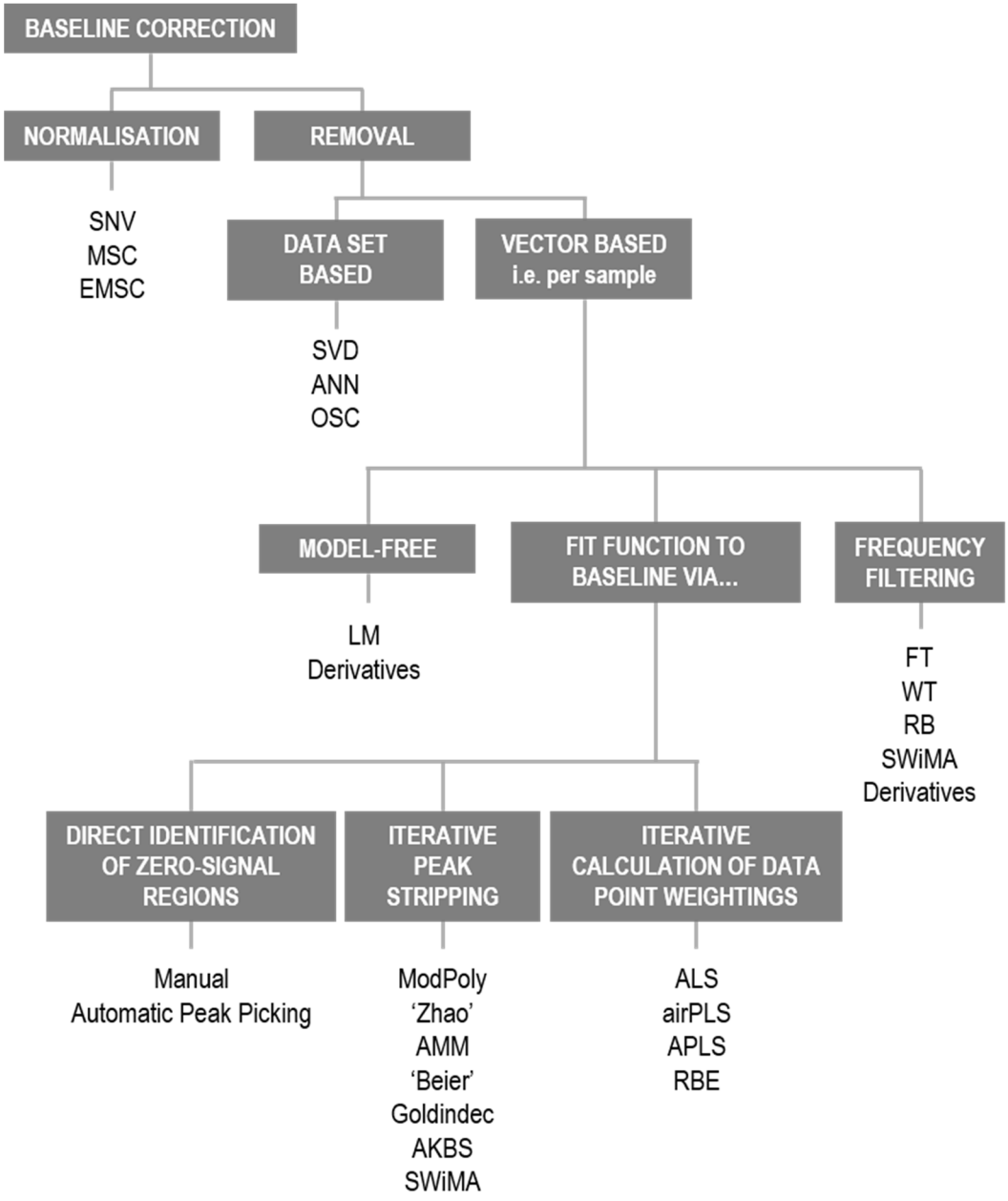


Figure 5-2 Framework and overview of some approaches to baseline correction. Examples given include standard normal variate (SNV), multiplicative scatter correction (MSC), extended multiplicative scatter correction (EMSC), singular value decomposition (SVD), artificial neural networks (ANN), orthogonal signal correction (OSC), local medians (LM), Fourier transform (FT), wavelets transform (WT), rolling ball (RB), small window moving average (SWiMA), modified polynomial (ModPoly), adaptive minmax (AMM), adaptive knots B-spline (AKBS), asymmetric least squares (ALS), adaptive iteratively reweighted penalised least squares (airPLS), adaptive-weight penalised least squares (APLS), and robust baseline estimation (RBE).

Frequency filtering methods are based on the premise that a raw spectrum is comprised of a high-frequency noise component, a mid-frequency signal component, and a low-frequency baseline component. By applying frequency filters to the raw spectrum the signal component or the baseline contribution may be isolated. Frequency filtering techniques include Fourier transforms (FT) [120], wavelets transforms (WT) [121], rolling ball (RB) [122], small-window moving average (SWiMA) [123] and derivatives. The latter is included in this category as well as the model-free category, since derivatives are essentially a kind of high-pass filter and in cases where the original spectrum is reconstructed, they are not model-free.

Modeling a baseline accurately with a function such as a polynomial or a spline, requires that 'zero-signal' regions in the spectrum are identified as 'anchor' points through which to fit the function. These may be directly identified via manual selection [124] or automatic peak picking (which by default identifies signal-free regions) [125, 126]. Alternatively, an iterative peak-stripping approach may be applied whereby each iteration eliminates points lying above the current fit until only the baseline remains. A highly cited paper on this kind of approach was published in 2003, presenting a modified polynomial (ModPoly) method [127]. Several subsequent papers have reported small variations on this method, including an initial peak removal step ('Zhao') [128], fitting several constrained and unconstrained polynomial functions and selecting the best points from these (adaptive minmax – AMM) [129], incorporating a contribution for known backgrounds ('Beier') [130], and changing the cost function by which the polynomial fits are optimised (Goldindec) [131]. Minimising a cost function (also known as loss function) is the means by which a best fit is mathematically determined. Commonly, a least squares cost function is used in the problem of fitting a polynomial or a spline through given data points. By iteratively weighting the contribution of different points in a spectrum to this calculation, however, according to whether they are likely to be points on the baseline or points in a signal, comprises another approach to arriving at a baseline fit. Examples of this kind of method include asymmetric least squares (ALS) [132], adaptive iteratively reweighted penalised least squares (airPLS) [133], adaptive-weight penalised least squares (APLS) [134] and robust baseline estimation (RBE) [135].

Few papers in the literature provide a direct comparison of different baselining methods on Raman spectra. Afseth *et al*. [136] applied ModPoly, differentiation and a number of normalisation techniques to a variety of biological spectra including salmon oils, salmon meat and juice samples. They found that normalisation methods such as SNV, MSC and EMSC could not handle the baseline features present in these spectra. Heraud *et al.* [137]

also compared first and second derivative methods with these normalisation techniques (SNV, MSC and EMSC) when applied to Raman spectra of individual microalgal cells. They confirmed that baselining was a necessary preprocessing step and reported that the derivative methods outperformed the normalisation methods. Leger and Ryder [138] compared ModPoly and differentiation applied to Raman spectra of illegal narcotics mixed with several diluents. They found that neither of these baseline removal approaches outperformed the other and so recommended ModPoly on the basis that the original peak shapes were retained. Beattie and McGarvey [114] asserted that using singular value decomposition (SVD) to construct and remove baselines from Raman spectra of human tissue samples (of Bruch's membrane) gave better results than baselining using manual techniques, ModPoly, or normalisation methods (SNV, MSC). It should be noted that in this case, however, the samples investigated gave rise to highly similar spectra so data set-based analysis would have been appropriate. Furthermore, the SVD method itself requires significant manual intervention in its current form.

The example studies given above are limited in the number and type of baselining techniques compared, perhaps being defined by what methods were readily available in a particular software. A greater number and variety of methods were investigated by Schulze *et al*. [115] and Liland *et al.* [139]. Schulze *et al*. presented an in-depth investigation in which many of the methods in Figure 5-2 were discussed critically particularly with respect to their suitability for automated analysis. Although the paper is an excellent survey, Schulze *et al*. only apply the baselining techniques to synthetic spectra. Liland *et al.* proposed a general strategy for selecting spectral baseline correction methods. As part of the work a large number of baselining techniques, including ALS, RBE, WT, LM, ModPoly, peak detection and RB, were trialled and written into a freely available package called 'baseline' [140] that can be executed in the statistical software R. These methods were applied to Raman spectra of fish oils and MALDI-TOF (Matrix Assisted Laser Desorption/Ionisation-Time of Flight) spectra of milk mixtures. In general it was found that ALS, ModPoly and RB performed well on the Raman spectra. It was emphasised, however, that these kind of results are specific to spectral type, instrument settings and sample type. Even so, in light of the large number of baselining methods available, case studies where these are applied to real Raman spectra are useful in future studies to be able to select and compare a few techniques that are likely to perform well.

5.1.3 Selection of Methods for Comparison

For processing large numbers of Raman spectra of EBPR bacteria, it is desirable that the baselining method employed is automated (requiring no user intervention) or semi-automated (requiring minimal user-intervention to set parameters), and that it can handle the variation present between spectra.

Normalisation and data set-based methods are not investigated here on the basis they are not expected to be able to cope with the variation in Raman spectra present in these samples. Differentiation methods can be useful because they are unbiased, but they can also distort spectra, potentially making them harder to interpret. In addition, Schulze *et al*. [115] surmised that the difficulty of peak picking (in fitting derivative spectra prior to spectral reconstruction) meant that these kind of methods may not be ideal for automation. Thus differentiation methods are also not investigated here. Manual baselining is commonly used and may be extended to batch (and hence semi-automated) processing if all the spectra in the batch can be assumed to have the same baseline anchor points. The Raman spectra of EBPR bacteria, however, vary greatly. An approach to baselining that involves assuming zero-signal regions are the same for each spectrum would therefore not be reliable for EBPR bacteria. Furthermore, zero-signal regions are not always straight-forward to select purely by visual inspection.

The selected baselining methods investigated in this work, therefore, represent approaches based on model-free methods, frequency filtering, iterative peak stripping and iterative calculation of data point weightings. Detailed explanations of the employed techniques are given in the methodology section

5.2 METHODOLOGY

5.2.1 Spectroscopic Measurements

Raman spectra were acquired with the instrumentation and the set-up described in Chapter 4. Spectra were acquired with a 200 second 'extended scan' over the range 400 to 1800 cm^{-1}. With this setting, each spectrum took 12 minutes to acquire. Bacterial samples and polyphosphate solution samples were prepared according to the methods described in Chapter 4.

5.2.2 Data Sets

Three data sets were used to compare and evaluate the baselining methods. The first comprised four Raman spectra of individual EBPR bacteria. These four spectra were selected purposefully to represent a range of spectral shapes: three spectra respectively contained polyphosphate, PHA and glycogen character, and one spectrum contained none of these. The second data set comprised 158 Raman spectra of individual EBPR bacteria from four different SBR samples. From each of these four samples, 42 spectra were collected and in the compiled data set of 168 spectra, 10 were discarded due to poor signal-to-noise ratio. The remaining spectra represented a range of spectral shapes expected to be encountered when analysing EBPR bacteria. Classes were assigned according to which bacterial storage compound signatures were visibly present (or not present) in the spectra. The third data set comprised 18 spectra of aqueous solutions of high molecular weight polyphosphate. The spectra were acquired from three preparations of a concentration series that (excluding each blank water sample) ranged from 0.05 M to 1 M with respect to phosphate monomer.

5.2.3 Preprocessing Methods

Nine automated baselining methods were investigated in this work. They were applied to Raman spectra following an initial background subtraction routine which removed the contribution of known background components (CaF_2 for data sets 1 and 2, and water for data set 3). The parameters for each baselining method were optimised on data set 1 by visual inspection (examples given in Appendix 2) before comparing the methods with each other. The algorithms and parameters used are summarised in Table 5-1. They are explained in more detail in the following text.

Table 5-1 Summary of background and baseline estimation methods investigated in this work, together with the optimum parameters determined for each method as applied to the test data sets.

Abbreviation	Description	Parameters *	Reference
ALS	asymmetric least squares	p = 0.001, maximum iter = 100, smoothing parameter = 300,000	[132]
AMM	adaptive minmax (modified polynomial)	fluorescence-to-signal (FS) ratio thresholds = [0.1 0.2 0.25 0.3 10]	[129]
APLS	adaptive-weight penalised least squares	iter = 10, smoothing parameter = 500,000	[134]
Beier	modified polynomial incorporating 'contaminant' spectrum	deg = 6, iter = 50, roi = [1 1491] (i.e. all spectral data points), 'contaminant' spectrum = mean (n=32) CaF_2 spectrum	[130]
BKG	Background removal: removal of orthogonal substrate signal component	substrate signal = smoothed, mean (n= 32) CaF_2 spectrum for bacterial samples, and smoothed, mean (n=7) water spectrum for polyphosphate solutions	[141]
LM	local medians	w_m = 350	[117]
ModPoly	modified polynomial	deg = 6, maximum iter = 100, tol = 0.001	[127]
RB	rolling ball	w_m = 150, w_s = 150	[122]
RBE	robust baseline estimation	w_m = 400, tuning constant = 0.1	[135]
SWiMA	small-window moving average	maximum iter = 55	[123]

* deg = polynomial degree, iter = number of iterations, p = weighting of positive residuals, roi = region of interest, tol = tolerance of difference between iterations, w_m = window width for algorithm, w_s=window width for smoothing

Background Subtraction

The vector correction method described by Maquelin *et al.* [141] was used to subtract known background signal from the raw Raman spectra. Most of the samples analysed in this work comprised individual bacterial cells dried on a CaF_2 substrate. The 'known background signal' used in these cases, therefore, was an average of 32 CaF_2 spectra. This mean CaF_2 spectrum was smoothed with a Savitzky-Golay filter with a window width of 15 and a polynomial order of 3.

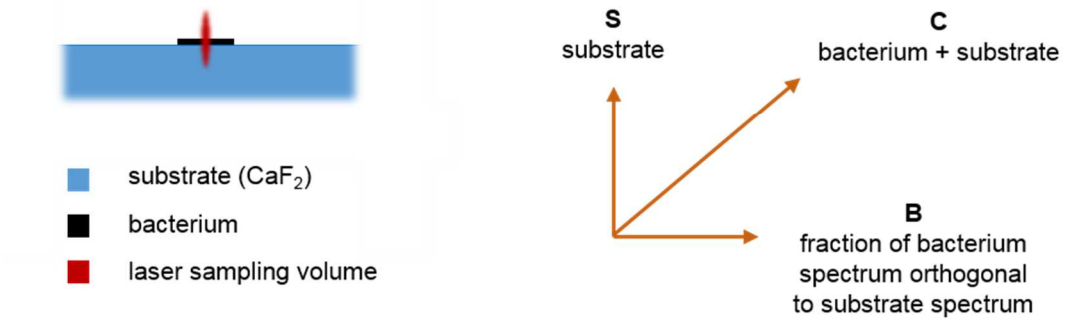


Figure 5-3 (Left) Diagrammatic representation (not to scale) of the sampling volume of a Raman instrument laser encompassing both the bacterium and the substrate. (Right) Depiction of the orthogonal vectors of the substrate **S** *and bacterium* **B** *spectra, and the combined substrate and bacterium spectrum* **C***. Figure adapted from [141].*

A diagrammatic representation of the background subtraction method is shown in Figure 5-3. The raw Raman spectrum is a combination of the signals arising from the bacterium and from the substrate. This can be represented by a vector **C** that can be decomposed into two orthogonal vectors: one measurable vector representing the pure substrate signal **S,** and the second vector, **B**, representing all the bacterial signal that is orthogonal to, and hence distinguishable from the substrate signal. To obtain **B**, a projection of **C** onto **S** is subtracted from **C** (Equation 5.1):

$$\mathbf{B} = \mathbf{C} - proj_S \mathbf{C} \qquad \textit{Equation 5.1}$$

The projection of a vector onto another vector, in this case **C** onto **S**, can be expressed in terms of dot products [142] (Equation 5.2):

$$proj_S \mathbf{C} = \frac{\mathbf{C} \cdot \mathbf{S}}{\|\mathbf{S}\|^2} \mathbf{S} \qquad \textit{Equation 5.2}$$

It follows that **B** can be obtained thus (Equation 5.3):

$$\mathbf{B} = \mathbf{C} - \frac{\mathbf{C} \cdot \mathbf{S}}{\|\mathbf{S}\|^2}\mathbf{S} \qquad \textit{Equation 5.3}$$

The mathematical routine is performed on first derivative spectra to negate the effect of any variance in fluorescence in the Raman spectra. As emphasised by Maquelin *et al.*, the result of the vector correction method does not represent the pure bacterium spectrum, but rather the component of the bacterium spectrum which is orthogonal to the substrate spectrum.

An in-house code was written to perform this method in MATLAB.

Iterative Peak Stripping

Three different polynomial-based iterative peak stripping methods were investigated in this work:

Lieber and Mahadevan-Jansen presented an automated, polynomial-based method for subtracting fluorescence from biological spectra [127]. This is an iterative process that progressively strips signal from the Raman spectrum so that the polynomial function is eventually fit to the baseline rather than the whole spectrum. This is achieved by cycling between least-squares fitting a polynomial function to the input spectrum, then producing an output spectrum that has selected the lowest intensity (from the raw spectrum or polynomial fit) for each data point (Figure 5-4). The process can be stopped either when a specified number of iterations has been completed (typically between 20 and 200), or when there is convergence in the number of data points reassigned in each iteration. This baselining method, referred to as 'ModPoly', was performed in R 3.2.2 using the package 'baseline' [140]. A polynomial order of 6 was used and a maximum of 100 iterations were run for each spectrum.

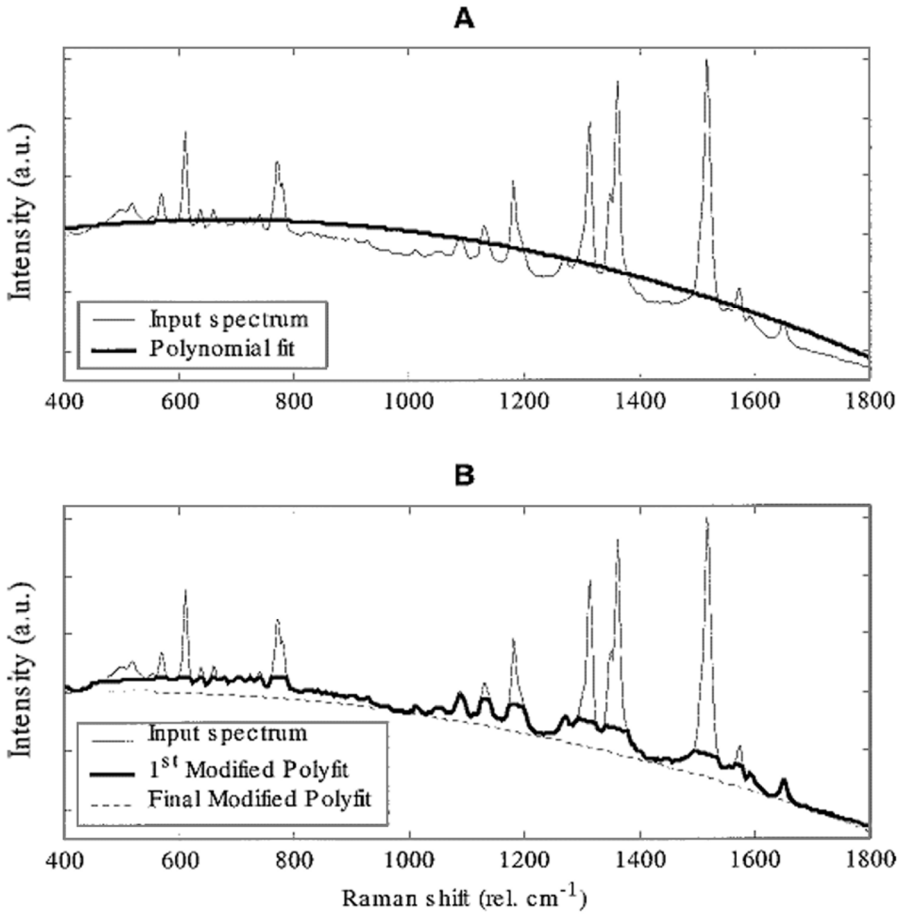


Figure 5-4 Polynomial curve fitting as it is applied to rhodamine 6G. (A): Raw measured Raman spectrum and a least-squared polynomial curve fit of the baseline fluorescence. (B): Raw measured Raman spectrum and two iterations of the modified curve fit applied until a smoothed polynomial curve along the base of the measured spectrum is obtained. Intensity is reported in arbitrary units (a.u.) Reproduced from [127].

Cao *et al.* [129] demonstrated how polynomial baseline fits are often poor at the end of a spectrum. If the endpoints are constrained to match the raw spectrum, however, the polynomial baseline fit can become distorted. Furthermore, the authors recognised that fluorescence-to-signal (FS) ratio of a spectrum affects which polynomial order gives the best baseline fit. In light of these facts, Cao *et al*. presented the adaptive minmax (AMM) method for automated baselining. In the first step, four different baseline fits – constrained and unconstrained fits for two different polynomial orders – are produced using the Lieber and Mahadevan-Jansen method. The polynomial orders selected are chosen automatically based on the FS ratio of the spectrum. The second step of the method selects the highest intensity at

each data point from each of these four fits to produce the final baseline fit. This routine was performed in MATLAB using code supplied by the authors in a personal communication. Subsequently, the code was modified for our spectra to force automatic selection of higher order (5th, 6th) polynomials (achieved by changing FS ratio thresholds for selecting polynomial order). In addition, normalisation and smoothing steps included within the routine were removed to make the method more comparable with other methods tested here.

Beier and Berger [130] recognised that the background of a Raman spectrum sometimes contains sharper features arising from known contaminants such as glass or water. These features cannot be fit well with a polynomial function so the authors modified Lieber and Mahadevan-Jansen's method to include a background reference spectrum. The process works to find the optimised weighting of background reference spectrum and polynomial function that best fits the baseline. This routine, referred to as 'Beier', was performed in MATLAB using code made freely available online [143] by the authors. A polynomial order of 6 was selected for the process and 50 iterations were run for each spectrum. The reference background spectrum used was an average of 32 CaF_2 spectra, smoothed with a Savitzky-Golay filter (window width=15, polynomial order=3).

Frequency Filtering

Two different baselining methods using principles of frequency filtering were investigated:

Kneen and Annegarn [122] proposed the rolling ball (RB) technique for baselining X-ray spectra, which is the mathematical equivalent of rolling a ball along the underside of a spectrum. The baseline is constructed by marking out the position of the top of the ball at each data point. Ideally, the diameter of the ball is selected to be greater than the width of the signal peaks, but less than the radius of curvature of the baseline. In this way, this process acts like a high-pass filter, effectively tracing out a low frequency baseline that may be subtracted from the spectrum. This procedure was performed in R 3.2.2 using the package 'baseline'. It should be noted that in this package, the Kneen and Annegarn method was simplified by not increasing the diameter of the ball as it moves across the spectrum [139] (a variable included originally to compensate for energy-induced peak-width broadening in X-ray spectra). Both the window width (i.e. diameter of the ball) and the smoothing window width used in this study was 150 pixels.

Schulze *et al*. [123] presented the Small-Window Moving Average (SWiMA) method for estimating baselines, which is based on the use of a zero-order Savitzky-Golay filter. This is a low-pass filter that is commonly used to reject high-frequency noise leaving a smoothed

spectrum. A characteristic of the filter is that as window width is increased, the frequency threshold is lowered so that more ‘noise’ is rejected and signals become more distorted. SWiMA is an iterative process that starts with applying the smallest (three point) window Savitzky-Golay filter to the raw spectrum. The raw and filtered spectra are then compared and the lowest intensity data point from each pair of data points is selected to construct the output spectrum. During each subsequent iteration the filter window width is increased by two points and the procedure is repeated. The result is that peaks are progressively stripped from the spectrum until either a specified number of iterations has been performed, or a stopping criterion has been reached at the point when just a baseline estimation remains. This routine was performed in MATLAB using code supplied by the authors in a personal communication. A maximum of 55 iterations were performed on each spectrum.

Model-Free Method

One model-free baselining method was investigated in this work:

Friedrichs [117] proposed the local medians (LM) approach for baselining NMR spectra. This works on the assumption that within a specified window width, the number of noise extrema is large relative to the number of true signal extrema. Note that a data point is an extremum if both its neighbours are larger or smaller than itself. The baseline points can then be initially approximated using the median intensity of all the extrema intensities within the window. A final estimated baseline is produced after applying a Gaussian function to smooth any sharp discontinuities. This procedure was performed in R 3.2.2 using the package ‘baseline’. It should be noted that their algorithm added some modifications to the procedure in order to be more applicable to types of spectra other than NMR [139]. The window width used in this study was 350 pixels.

Iterative Calculation of Data Point Weightings

Three baselining methods using some kind of weighted least squares fitting were investigated:

Eilers and Boelens [132] described the asymmetric least squares (ALS) approach to baselining. This method iteratively fits a spectrum with a Whittaker smoother [144, 145] to calculate a baseline. For the first fit, the weightings are all set to 1 and an initial baseline is estimated. In subsequent iterations, positive residuals (variables above the baseline) are down-weighted because negative residuals are more likely to be important for fitting the baseline. Two parameters are necessary for the algorithm, the first being the magnitude of

the weighting, p. The weightings of positive residuals are then equal to p and the weightings of negative residuals are set to (1-p). The second parameter required is a smoothing weighting which tunes the trade-off between fidelity of a baseline to the data and the smoothness of the estimated baseline. With p tending to zero and a large smoothing weighting, the estimated baseline will tend to follow the valleys of the original spectrum. This procedure was performed in MATLAB using PLS_Toolbox. The weighting magnitude (p) was set to 0.001 and the smoothing parameter was 300,000. A maximum of 100 iterations were used for each spectrum.

Cadusch *et al.* [134] presented the adaptive-weight penalized least squares (APLS) algorithm for baseline correction. The procedure starts with an initial baseline estimate for which the mean of the observed signal can be used. For each observed point in the spectrum, a Poisson probability is then calculated – according to the likelihood of that point belonging to baseline or signal given the current estimated baseline point. The calculated probabilities are used as the weightings in the subsequent penalized least squares [133] fit– where a new baseline is estimated by minimising a spline-type cost function. The algorithm then continues to cycle between calculating new weightings (based on Poisson probabilities) and new baseline estimates (based on minimising a cost function) until there is little change in the calculated weightings, or for a user-defined number of iterations. Only one given parameter is necessary in the algorithm – a smoothing weighting that tunes the trade-off between fidelity of a baseline to the data and the smoothness of the estimated baseline. Currently there is no automated way to optimise this parameter, but rather it is determined by trial and error. Contrary to methods that use a binary weighting of peaks and baseline [127, 132, 133], the APLS procedure calculates continuously weighted variables that are asserted to improve the baseline estimation. This routine was performed in MATLAB using code supplied by the authors in a personal communication. A smoothing parameter of 500,000 and a maximum of 10 iterations were used for each spectrum.

Ruckstuhl *et al.* [135] introduced a technique called robust baseline estimation (RBE) which is closely related to the well-established “locally weighted scatterplot smoother” (LOWESS) algorithm [146]. LOWESS employs the principals of both local regression and robust estimation to fit data in the case that a global functional form is either unknown or undesirable. With local regression, the data model is built from fitting smaller subsets of the data with a low order polynomial (typically linear) function. With robust estimation, the contributions of the data points to the least-squares fit are weighted with ‘robustness weights’ so that outliers (extreme observations) have much less influence. The LOWESS

procedure is iterative. The cycle alternates between localised regression and weights calculation (which are incorporated into the subsequent regression) until convergence is achieved (typically in 3-5 iterations). RBE is very similar to LOWESS, but differs in that the calculated weights are asymmetric to account for the fact that in a spectral baselining problem, the outliers in the spectrum are typically unidirectional. Negative residuals are assigned a weighting of 1, whereas the weightings of positive residuals are calculated according to Tukey's bisquare weights. The two key parameters in RBE are the bandwidth (or window width) that defines how many data points are included in localised regression, and a tuning constant that specifies how sensitive the fit is to outliers. RBE was performed in R 3.2.2 using the package 'baseline'. A window width of 400 and a tuning constant of 0.1 was used in the given example.

5.2.4 Evaluation of Baselining Methods

The success of a baselining technique ultimately depends on the purpose of the spectra acquired. The evaluation of baselining methods presented here addressed three potential purposes – aesthetic appeal, classification and quantification. First, the effect of baselining on a small set of spectra (data set 1) were compared by inspection. Second, two classification techniques – Multiplicative Curve Resolution (MCR) and Partial Least Squares Discriminant Analysis (PLS-DA) – were applied to a set of 158 baselined EBPR bacteria spectra (data set 2). With as few assumptions as possible, MCR attempts to determine pure component spectra and related concentrations that together compose the overall spectra. By contrast, PLS-DA builds a model based on PLS regression that predicts the class to which a sample belongs. Third, potential effects on quantification were assessed. Unfortunately, no concentration data was available for individual EBPR bacteria due to the lack of methods to measure parameters such as polyphosphate, glycogen or PHA concentration in single cells. Instead, PLS regression was applied to Raman spectra of varying concentrations of a polyphosphate solution (data set 3), and the model statistics compared for different baselining methods.

5.3 RESULTS AND DISCUSSION

5.3.1 Appearance of Corrected Spectra

The top-left spectrum ('BKG') in Figure 5-5 shows the result of background subtraction (vector correction method described in 5.2.3) applied to a raw EBPR bacterium spectrum. From the mean CaF_2 (shown in green) and calculated background (shown in orange) included in the figure, it is clear that the exact shape and intensity of the input CaF_2 spectrum is not subtracted from the EBPR bacterium spectrum by this method. The background component identified and removed does, however, strongly reflect the character of the CaF_2 spectrum. The background-corrected spectrum has a noticeable change in relative peak intensities between 1200 and 1500 cm^{-1}, corresponding to the region containing the sharpest feature in the CaF_2 spectrum. In addition, the corrected spectrum is more concave than the raw spectrum in the region 800 to 1500 cm^{-1} with respect to the zero intensity line.

The remaining spectra in Figure 5-5 show the results of nine different automated baselining methods applied to the background-corrected EBPR bacterium spectrum. Four of the calculated baselines show relatively poor fits to the spectral data: rolling ball ('RB'), local medians ('LM'), robust local regression ('RBE') and a modified polynomial method that incorporates a background ('Beier'). The rolling ball corrected spectrum exhibits some negative regions, particularly around 800 and 1500 cm^{-1}. In addition, the baseline deviates significantly from the spectrum at the lowest wavenumbers. During parameter optimisation, no trialled window width (i.e. 'ball diameter') seemed to give a good baseline fit, providing sufficient resolution and yet not causing peak erosion in the corrected spectrum. The method appears to be unsuitable for this type of spectrum which contains broad, overlapping peaks. The same is true for the local medians method where the calculated baseline is weighted much too positively by the broad spectral peaks. The result is that the corrected spectrum is negatively offset and has a poorly fitting baseline. The rolling ball and local medians methods were designed originally for NMR and X-ray spectra respectively, both of which have much narrower peaks. It is perhaps no surprise then, that with no further modifications they did not provide a well-fitting baseline for this EBPR bacterium Raman spectrum.

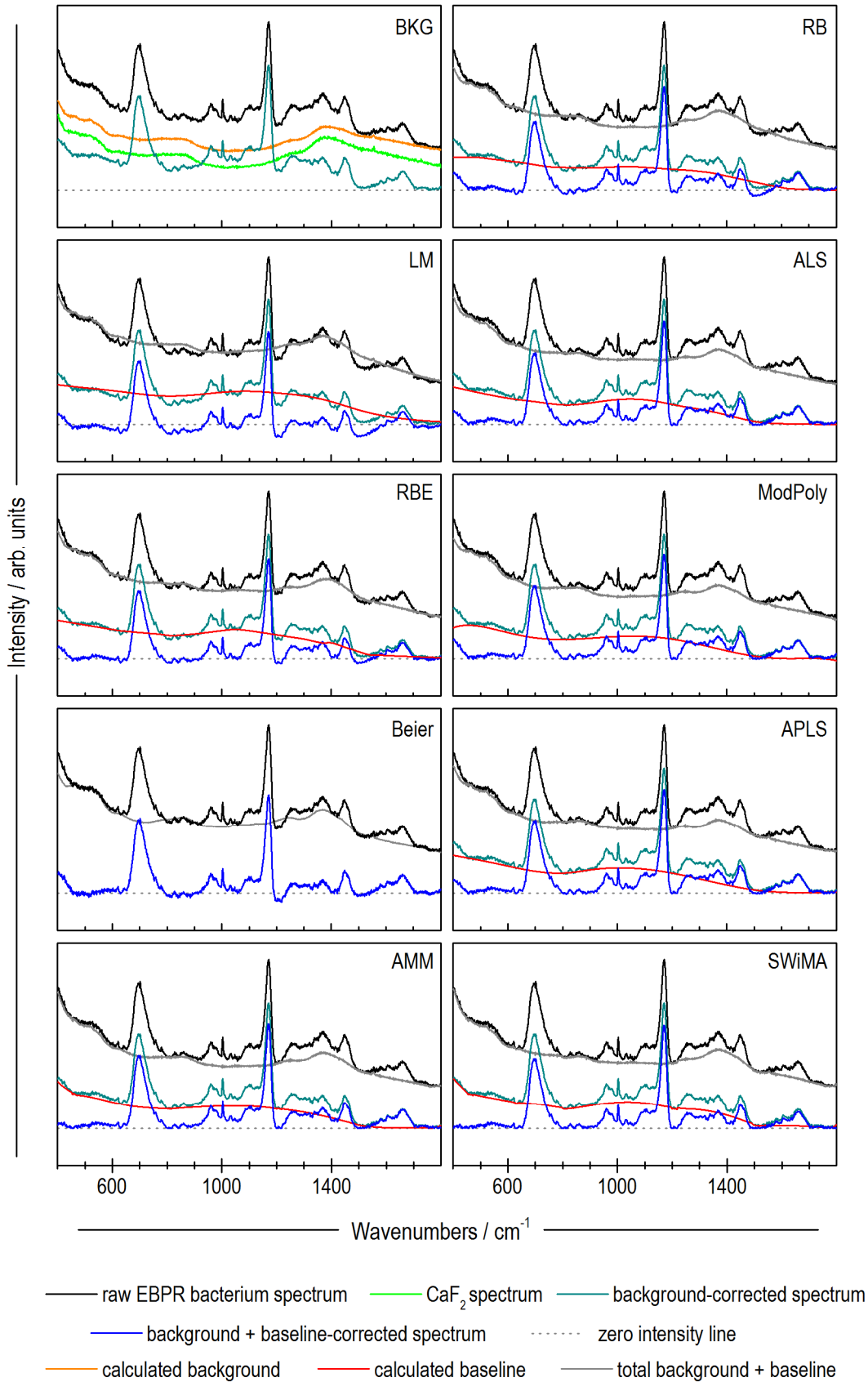

BKG
RB
LM
ALS
RBE
ModPoly
Beier
APLS
AMM
SWiMA
Intensity / arb. units
600
1000
1400
Wavenumbers / cm-1
raw EBPR bacterium spectrum
CaF2 spectrum
background-corrected spectrum
background + baseline-corrected spectrum
zero intensity line
calculated background
calculated baseline
total background + baseline

Figure 5-5 One background subtraction and nine baselining methods demonstrated on a single EBPR bacterium Raman spectrum. The methods and parameters are summarised in Table 5-1 and are abbreviated as follows: BKG – background subtraction (vector correction method), RB – rolling ball, LM – local medians, ALS – asymmetric least squares, RBE – robust baseline estimation, ModPoly – modified polynomial, Beier – modified polynomial incorporating a background spectrum, APLS – adaptive-weight penalised least squares, AMM – adaptive minmax, SWiMA – small-window moving average.

The RBE-corrected spectrum is similar to the RB-corrected spectrum, albeit with a more negative region around 1200 cm^{-1} and a tighter-fitting baseline around 1050 cm^{-1}. The authors recommended using a 'robustness' tuning constant of 3 for the algorithm, however, with this setting the baseline was much too positively offset, similar to that shown for LM. Instead, we reduced the tuning constant to 0.1 to make the technique less sensitive to outliers but even with this extremely low value the calculated baseline was still too positive in some regions.

The modified polynomial method incorporating a background ('Beier') was applied directly to the raw EBPR bacterium spectrum (i.e. without prior background subtraction) since this baselining algorithm itself included background subtraction. The method iteratively calculates a baseline using weighted contributions from the background spectrum and a polynomial fit. In this case the calculated baseline was found to fit the EBPR bacterium spectrum poorly. One reason for this could be if the mean (n=32) CaF_2 spectrum used in the algorithm as the background spectrum, was not similar enough to the true CaF_2 contributions present in this particular sample. This is possible since the broad features of the CaF_2 spectrum may also be subject to fluorescence and hence the spectral shape may vary slightly between samples and measurements. The requirement to measure a background spectrum for every individual sample would double the acquisition time of a sample set. This method, therefore, does not seem to be the best available for these particular spectra. Note that in contrast to this algorithm, the orthogonal vector correction method for background subtraction copes with any variation in fluorescence of the background and sample spectra by performing the algorithm on first derivative spectra.

Five of the calculated baselines in Figure 5-5 show relatively good fits to the spectral data: asymmetric least squares ('ALS'), adaptive minmax ('AMM'), adaptive-weight penalised least squares ('APLS'), modified polynomial ('ModPoly'), and small-window moving

average ('SWiMA'). None of the spectra corrected with these methods show obvious peak erosion (overfitting) or regions of negatively offset baseline (underfitting). A more direct comparison of these baselining algorithms is demonstrated on four different background-corrected EBPR bacteria Raman spectra in Figure 5-6.

Four of the baselining methods – APLS, ALS, ModPoly and AMM yield relatively smooth baselines whereas SWiMA tends to give a baseline that is smooth in regions but has some sharper features joining these regions. This ability for SWiMA to produce this cusp-type shape means that whatever the 'tuning' of the fit (in this case determined by how many iterations are performed), adjacent regions in the spectra will have a similar closeness of fit. By contrast, if a polynomial-type function is tuned (by the polynomial degree) to fit more closely to the spectra in one region, the adjacent region is likely to be underfit. This can be seen in the ModPoly baselines for the region 800 to 1200 cm^{-1} compared with the region 1200 to 1500 cm^{-1}. The ability of SWiMA to fit adjacent regions with similar fit 'closeness' is an advantage, however, ideally the baselines would be a little smoother and would not cut sharply into any peak regions. ALS baselines give a similar closeness of fit between regions without the sharp features present in the SWiMA baselines.

In all four spectra in Figure 5-6, SWiMA and AMM yield a well-fitting baseline at the spectral extremes (i.e. highest and lowest wavenumbers). The baselines calculated by ModPoly, ALS and APLS in these regions, however, are seen to deviate from the spectra in one or more cases. For many spectra, the extreme regions don't contain important information. For the EBPR bacteria Raman spectra acquired here, however, some peaks of interest are present close to the extremes – such as the PHA peaks at ~430 cm^{-1} and ~1725 cm^{-1}. It is therefore important that the baseline fits well at the extremes. In future, acquiring spectra with a slightly wider wavenumber range would mean that these fits are less important.

A large difference in calculated baselines is consistently seen in the region 1200 to 1500 cm^{-1}. This region encompasses peaks which, in biological spectra, are typically assigned to amide III and CH_2 deformation bands. In all four spectra, SWiMA gives the most positive estimation of the baseline in this region, consequently yielding corrected spectra with the lowest peak intensities. Conversely, three methods – ModPoly, APLS and AMM – give the most negative estimation of baseline in this region resulting in corrected spectra with the highest peak intensities. The ALS method tends to calculate a baseline that lies between these two extremes.

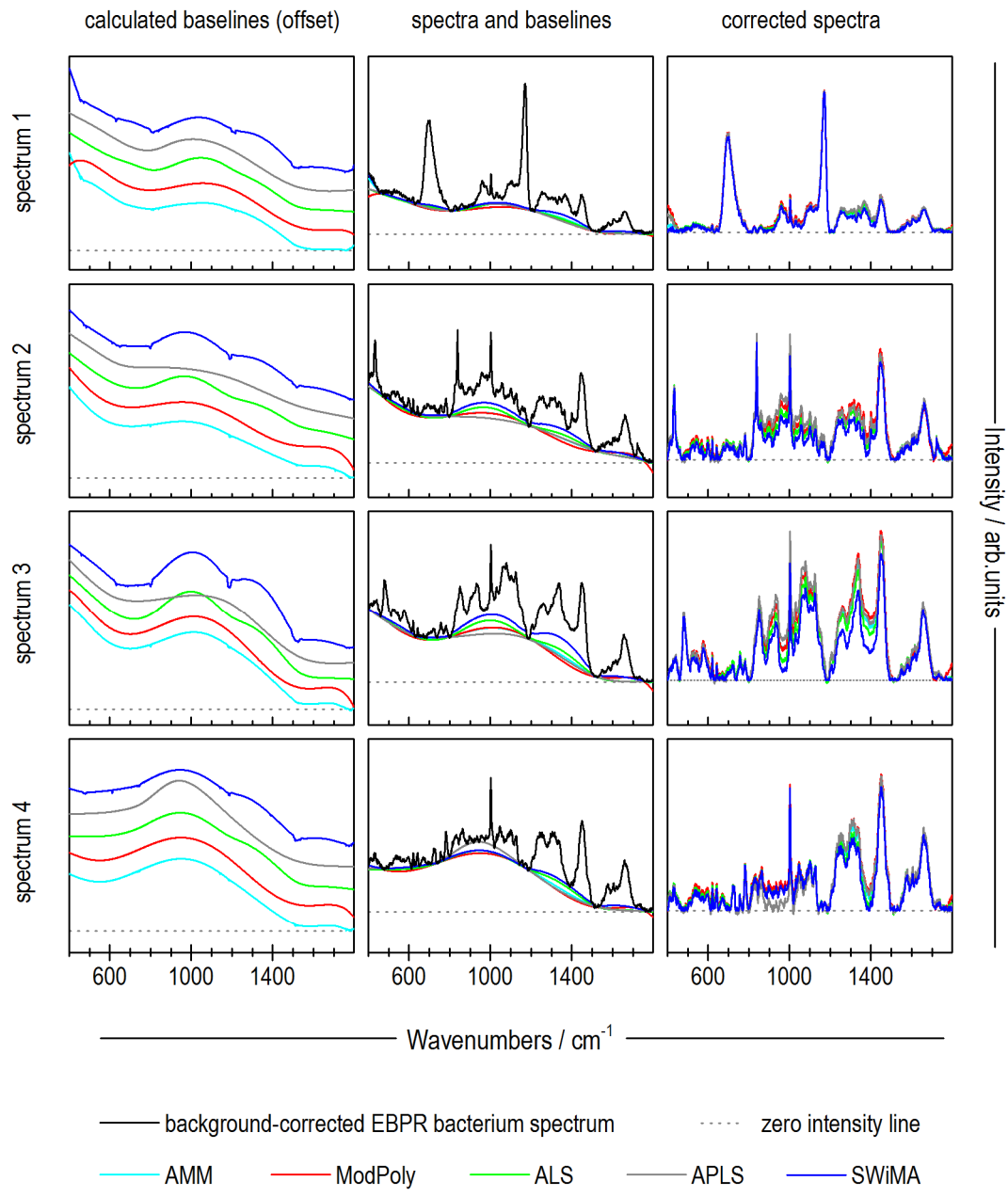


Figure 5-6 Five different baselining methods demonstrated on four different background-corrected EBPR bacterial Raman spectra (numbered 1 to 4). The calculated baselines are shown on the left, offset for clarity. These baselines together with the original spectra are shown in the centre, and the background-plus-baseline-corrected spectra are shown on the right. The baselining methods and parameters are summarised in Table 5-1 and are abbreviated as follows: ALS – asymmetric least squares, ModPoly – modified polynomial, APLS – adaptive-weight penalised least squares, AMM – adaptive minmax, SWiMA – small-window moving average.

In the region 800 to 1200 cm^{-1}, that includes the well-known phenylalanine peak near 1002 cm^{-1}, a similar trend is followed. Spectra corrected by ModPoly, APLS and AMM have the highest peak intensities in most cases. SWiMA generally results in the lowest peak intensity, and again, ALS-corrected peak intensities lie between these two extremes. In spectrum 4 a deviation from this trend is seen, where APLS fits the baseline in this region much more closely than the other methods.

For our EBPR bacteria spectra, it should be noted that no pure spectrum is available with which to compare the baseline-corrected spectra. It is impossible, therefore, to say which baselining method yields the most 'correct' result. For classification and quantification purposes, it is not obtaining an exact, true spectrum that matters the most, but rather the consistency of a method and its effectiveness in removing unwanted variation between samples.

5.3.2 Potential Effect on Classification

Both MCR and PLS-DA were applied to a data set comprising 158 Raman spectra of individual EBPR bacteria to observe potential effects that different baselining methods could have on classification-type analyses. MCR loads are examined since this kind of model exclusively yields positive loadings which are intended to represent pure components. They are typically easier to interpret than loadings calculated by PCA or PLS. The results of MCR analysis are shown in Figure 5-7. Seven different preprocessing methods were compared: no preprocessing (Raw), background subtraction (BKG), and background subtraction followed by five different baselining methods – ALS, AMM, APLS, ModPoly and SWiMA – which were selected on the basis that they performed well in a visual comparison of methods (Figure 5-5). No further preprocessing such as normalisation was applied. The MCR models were built with 4 components to represent a fundamental bacterium spectrum plus the three known storage compounds (polyphosphate, PHA, glycogen) that can dominate a Raman spectrum when present.

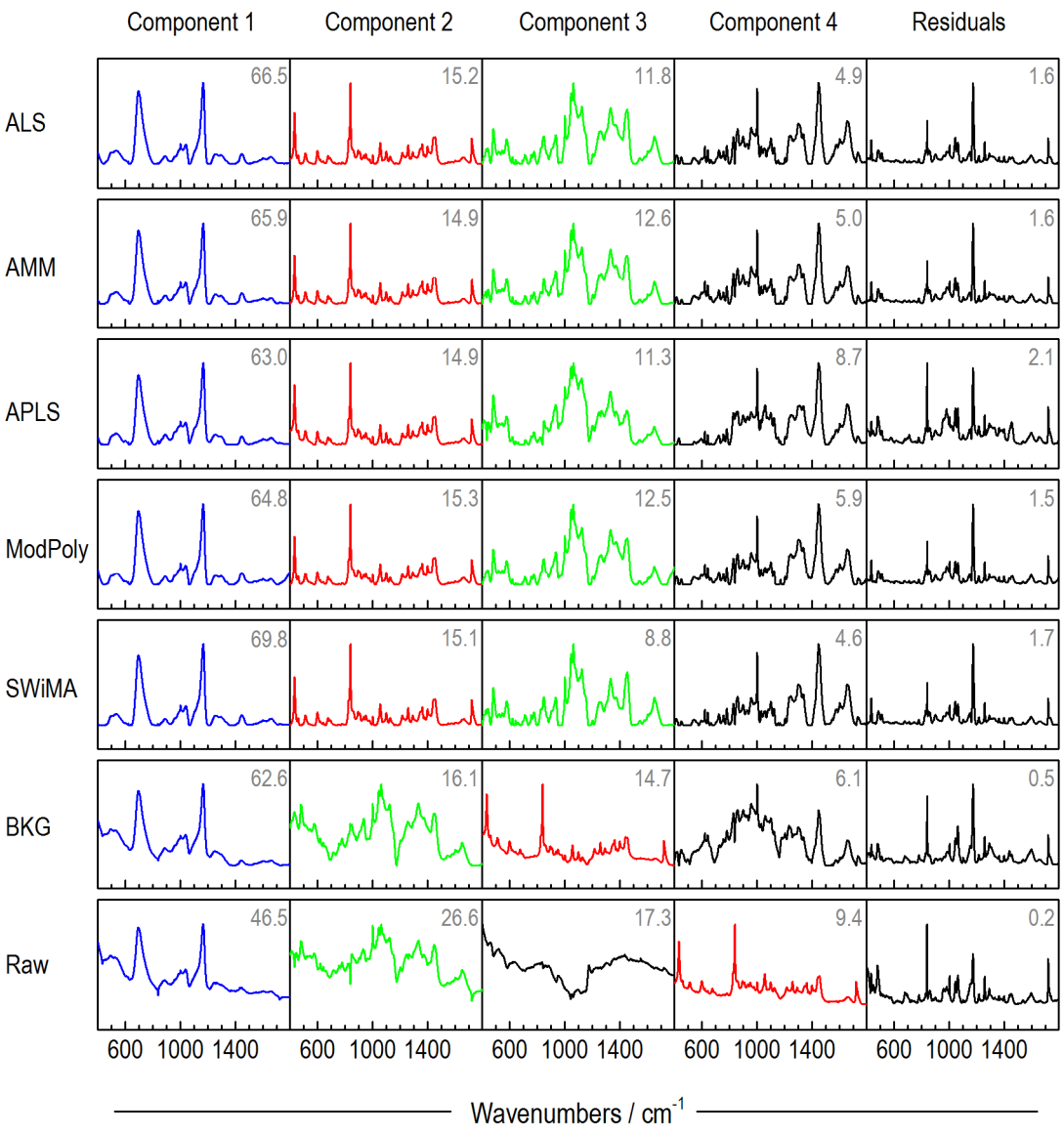


Figure 5-7 Loadings for 4-component MCR models built on a test set of 158 Raman spectra of individual EBPR bacteria. The Raman spectra were preprocessed seven different ways: no preprocessing (Raw), background subtraction (BKG), and background subtraction followed by five different baselining methods – asymmetric least squares (ALS), adaptive minmax (AMM), adaptive-weight penalised least squares (APLS), modified polynomial (ModPoly), and small-window moving average (SWiMA). No further preprocessing such as normalisation was applied. Percentage variance captured by each component is indicated in grey text and the components are ordered in decreasing value of variance captured. Loadings strongly reflecting the character of polyphosphate, PHA and glycogen Raman spectra are shown in blue, red and green respectively.

For all five of the baselining methods applied, Figure 5-7 shows that three of the four MCR model components are similar to the pure compound spectra for polyphosphate, PHA and glycogen. These are shown in blue, red and green respectively.

The MCR models built on raw and background-only-corrected spectra also yield components that contain clear contributions from polyphosphate, PHA and glycogen. However, the presence of fluorescence signal in each of these components is also apparent. In addition, the third component of the model built on raw spectra is not recognisable as either a storage compound or a typical bacterium spectrum.

For ease of comparison, Figure 5-8 shows three of the four MCR model components (for all seven models) overlaid with the pure compound spectra they most closely resemble.

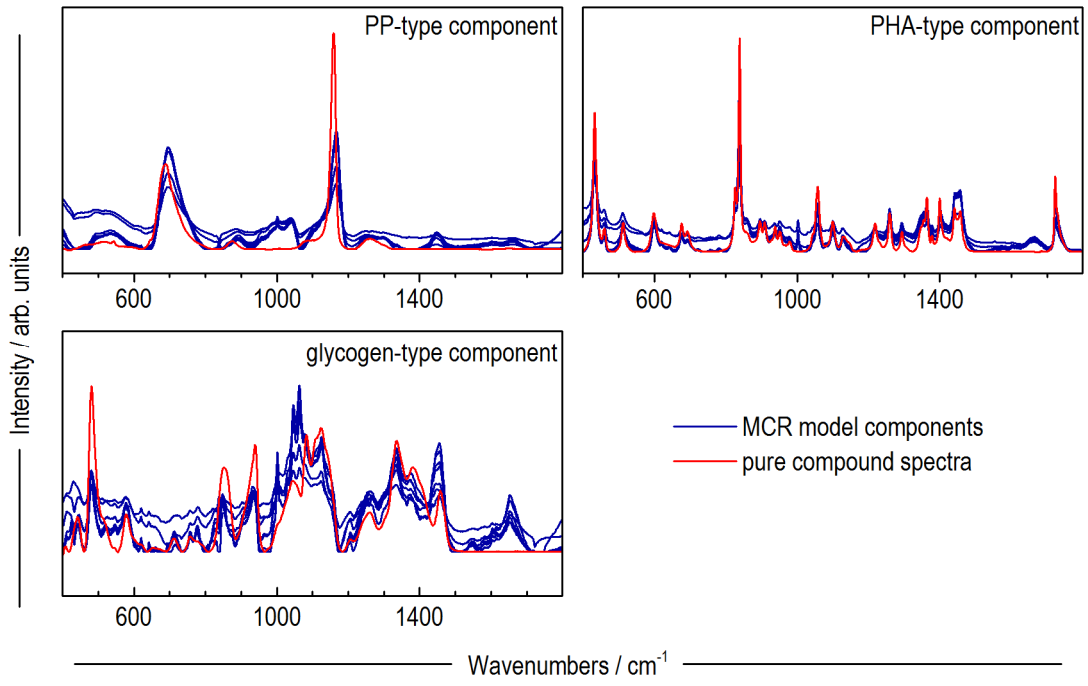


Figure 5-8 Comparison of MCR model components (in blue) and pure compound spectra (in red). The MCR model components are derived from Raman spectra preprocessed in all seven different ways discussed in the text. These are overlaid with spectra of the pure compounds polyphosphate (PP), polyhydroxyalkanoate (PHA) and glycogen. The pure compound spectra were baselined manually.

Based on the observations above, if the purpose of an MCR model is simply to determine the presence or absence of a particular component – in this case to determine the presence or absence of a particular metabolite in a set of unknown spectra – then the inclusion of background or fluorescence signal in the components may have little effect on the analysis. If, however, the purpose of the model is to determine the relative quantities of each component present in a set of unknown spectra, then it can be expected that the inclusion of background or fluorescence signal in the components will introduce error into the scores calculated for each component. This is because the background (substrate) and fluorescence signal contributions to the spectra are not constant between samples.

To illustrate this point, Figure 5-9 shows 32 Raman spectra of a calcium fluoride substrate, all measured with the same acquisition time and the laser spot focused manually onto the surface of the material. The variation in spectral intensity between these measurements is evident and even when normalised to unit area, the spectra do not overlie one another exactly. Reasons for the observed differences are likely to include slight changes in laser focal position, instrument alignment, CCD sensitivity, and laser power and wavelength over time. Thus, even if a background (substrate) spectrum were recorded for each sample spectrum in order to do a direct subtraction of the background contribution, this would not necessarily work well. Furthermore, it effectively doubles the acquisition time since two spectra would be required per sample. It was for these reasons that the vector correction method described by Maquelin *et al.* (previously discussed) rather than direct subtraction was used to remove background spectra from the raw Raman spectra.

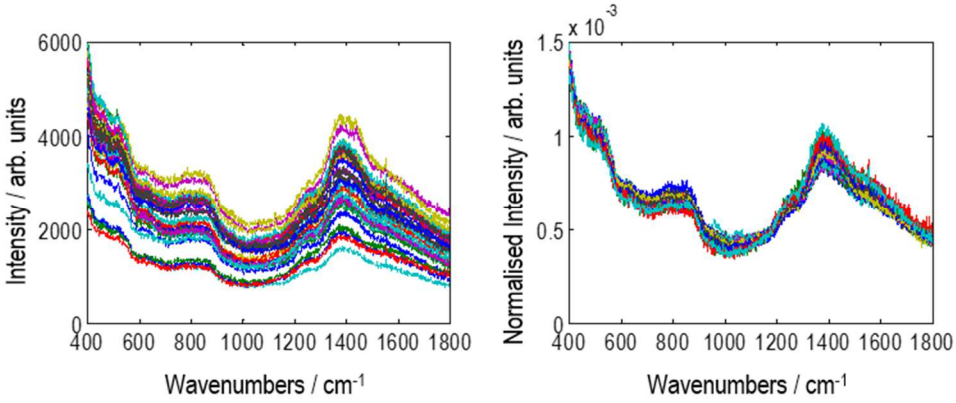


Figure 5-9 32 Raman spectra of a calcium fluoride substrate (left) and the same spectra normalised to unit area (right).

The results of seven PLS-DA models built on the 158 spectra are shown in Figure 5-10. As for MCR analysis, each model represents one of seven different ways of treating the spectra with respect to background and baselining. Before analysis, each spectrum was assigned one of five classes according to which metabolite signals were visibly present. By inspection, 76 spectra within the data set (48 %) contained polyphosphate, 26 (16 %) contained PHA, 24 (15 %) contained both polyphosphate and PHA, 15 (10 %) contained glycogen, and 17 (11 %) contained no visible metabolites. Further preprocessing comprised normalising each spectrum to its most intense peak so that each sample had a similar influence on the model. In addition, the spectra were mean centred. The number of latent variables (LVs) used to build each model was selected based on the corresponding variance in mean classification error (Figure 5-10) of the cross validation (venetian blinds, 10 splits) results. The number of variables used in each model, mean classification error, and spectral variance captured are summarised in Table 5-2.

Table 5-2 Summary of selected PLS-DA model parameters and statistics

PLS-DA Model	Number of Latent Variables	Mean Classification Error (for cross-validation)	% Spectral Variance Captured
ALS	5	0.052	94.6
AMM	5	0.052	94.8
APLS	6	0.060	94.7
ModPoly	5	0.052	94.8
SWiMA	5	0.058	94.2
BKG	5	0.066	95.8
Raw	7	0.066	98.3

Five of the models were built with 5 LVs, but APLS and Raw models were built with 6 and 7 LVs respectively. These selections were based on where the curve in the mean classification error started to flatten out, and thus for APLS and Raw, this point was observed at a greater number of LVs than for the other of models. The five models built on background and baselined spectra (ALS, AMM, APLS, ModPoly and SWiMA) showed very similar model statistics – with mean classification errors ranging from 0.052 to 0.060 and

percentage spectral variance captured ranging from 94.2 to 94.8%. The Raw PLS-DA model captured the greatest percentage spectral variance (98.3%) and yet had the joint highest mean classification error (0.066) with the BKG model.

Each separate graph in Figure 5-10 shows the sensitivity and specificity for PLS-DA based class assignment according to the cross validation results. Sensitivity is the number of samples predicted to be in a class divided by the number of samples actually in the class, otherwise known as the true positive rate. Specificity is the number of samples predicted to not be in a class divided by the number of samples actually not in a class, also known as the true negative rate. Reporting cross validation results rather than calibration model results gives a predictive estimate of sensitivity and specificity for a larger, similar set of samples to those on which each model was built.

Overall, the results of PLS-DA analysis are very similar for all seven models. Sensitivity and specificity for each class is discussed in more detail below although it should be noted that the scale of the differences discussed is very small in most cases.

With respect to the GLY class, ALS, AMM and ModPoly models had the highest sensitivity although the difference over the other models corresponds to just one more sample being correctly classified as containing glycogen. All the models had the same specificity except for the Raw model for which it was slightly lower, corresponding to one extra sample being wrongly identified as not belonging to the glycogen class.

With respect to the PHA class, all the models which were background and baseline corrected (ALS, AMM, APLS, ModPoly and SWiMA) had the same sensitivity and specificity, both of which were greater than those for the other two models. The difference between greatest and worst sensitivity and specificity both corresponded to the classification of two samples.

Surprisingly, the Raw model had the greatest sensitivity toward the PP class. The BKG and ALS models had the worst sensitivity toward this class, classifying two less spectra correctly than the Raw model. With respect to specificity, all the models except for APLS correctly predicted 77 out of 82 spectra as not belonging to the PP class. APLS had a slightly lower specificity, predicting 76 spectra did not belong to this class.

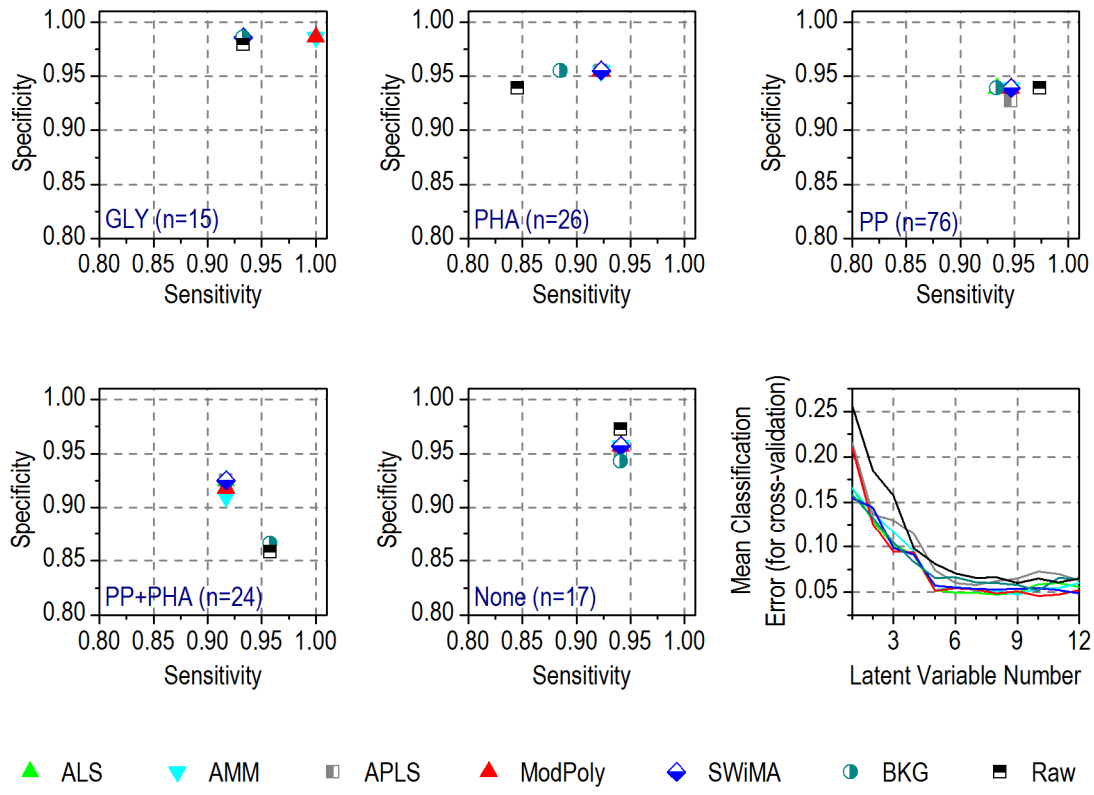


Figure 5-10 Depiction of the cross-validation specificity and sensitivity for classification of 158 Raman spectra of individual EBPR bacteria based on seven different PLS-DA models. Each model represents one of seven different ways of treating the spectra with respect to background and baselining: no preprocessing (Raw), background subtraction (BKG), and background subtraction followed by five different baselining methods – asymmetric least squares (ALS), adaptive minmax (AMM), adaptive-weight penalised least squares (APLS), modified polynomial (ModPoly), and small-window moving average (SWiMA). The number of latent variables (LVs) used to build each model was selected based on the corresponding variance in mean classification error (also shown, line colour corresponds with symbol colour). Five of the PLS-DA models were built with 5 LVs, the APLS model was built with 6 LVs, and the Raw model was built with 7 LVs. Prior to building the models, each spectrum was assigned to one of five classes based on which metabolite signals were visibly present: glycogen (GLY), polyhydroxyalkanoate (PHA), polyphosphate (PP), polyphosphate and polyhydroxyalkanoate (PP+PHA) and no metabolite (None). The number of spectra originally assigned to each class by inspection is indicated in brackets.

With respect to the PP+PHA class, the Raw and BKG models had the greatest sensitivity, correctly classifying one more spectrum as belonging to this class than the other five models. All these five background and baselined models had much greater specificity toward this class, however, correctly identifying between six and nine more spectra as not belonging to this class. Overall then, the background and baselined models performed better with respect to this class.

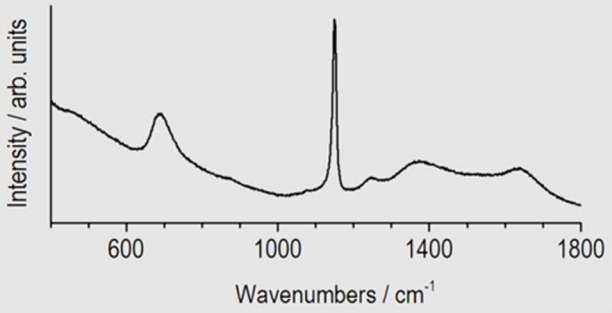
All seven models had the same sensitivity toward the None class, correctly identifying 16 out of 17 samples as belonging to this class. With respect to specificity, the Raw and BKG models gave the best and worst results respectively. The difference between these equated to 4 samples being correctly identified as not belonging to the None class. All background and baselined models (ALS, AMM, APLS, ModPoly and SWiMA) had the same specificity toward this class.

Ideally, classification results would have both high sensitivity and high specificity, so the best results are indicated by a position in the top-right of these plots. No one model was consistently positioned the furthest toward the top-right corner for all classes. When considering the results as a whole, therefore, it can be concluded that all the baselining methods trialled here were found to perform similarly well with respect to classifying EBPR spectra according to the metabolites they contained. On the other hand, the mean classification error results are higher for the Raw and BKG models (both 0.066) than the other five models (between 0.052 and 0.060) indicating that baselining the spectra as part of preprocessing can consistently improve the sensitivity and specificity of classification of the Raman spectra of EBPR bacteria with respect to metabolite content.

5.3.3 Potential Effect on Quantification of Polyphosphate

The model statistics from PLS regression analysis on three preparations of a concentration series of high molecular weight polyphosphate are shown in Table 5-3. The model statistics shown are mean values for the three preparations, with the standard deviations shown in italics. Again, seven different preprocessing methods were compared: no preprocessing (Raw), background subtraction (BKG), and background subtraction followed by five different baselining methods – ALS, AMM, APLS, ModPoly and SWiMA. The PLS models were built with one latent variable and no further preprocessing such as normalisation or mean centring was applied. ‘Leave one out’ cross validation was used to gain some idea of the predictive capacity of each model.

Table 5-3 Mean PLS regression model statistics for the Raman spectra of three concentration series of high molecular weight polyphosphate. Standard deviations are shown in brackets. A spectrum of 1 M (in phosphate monomer) polyphosphate is shown above the table. Results are shown for models built after seven different ways of preprocessing the data: no preprocessing (Raw), background subtraction (BKG), and background subtraction followed by five different baselining methods – asymmetric least squares (ALS), adaptive minmax (AMM), adaptive-weight penalised least squares (APLS), modified polynomial (ModPoly), and small-window moving average (SWiMA).



Baselining Method	PLS regression model statistics *			
	% X variance captured	% Y variance captured	RMSECV	R^2 (CV)
ALS	99.09 *(0.3)*	99.82 *(0.1)*	0.0344 *(0.015)*	0.9924 *(0.006)*
AMM	99.16 *(0.3)*	99.80 *(0.1)*	0.0360 *(0.018)*	0.9914 *(0.007)*
APLS	99.26 *(0.3)*	99.85 *(0.1)*	0.0330 *(0.016)*	0.9927 *(0.006)*
ModPoly	99.14 *(0.3)*	99.81 *(0.1)*	0.0353 *(0.014)*	0.9922 *(0.006)*
SWiMA	99.19 *(0.3)*	99.82 *(0.1)*	0.0350 *(0.017)*	0.9917 *(0.006)*
BKG	96.05 *(0.7)*	97.65 *(0.2)*	0.1134 *(0.031)*	0.9573 *(0.044)*
Raw	99.03 *(0.1)*	54.67 *(2.1)*	0.3991 *(0.040)*	0.7945 *(0.109)*

* PLS = partial least squares, X variance corresponds to spectral variance, Y variance corresponds to concentration variance, RMSECV = root mean square error of cross-validation, R^2 = correlation coefficient, CV = cross-validation

Percentage X and Y variance captured reflects how much spectral and concentration variance respectively is represented in the PLS model. The PLS algorithm aims to maximise covariance i.e. not just to capture as much spectral (X) variance as possible, but to find the variance in X that is most useful for predicting concentration (Y). Since the samples studied here are simple concentration series of aquous polyphosphate, it could be expected that one latent variable (component) would be sufficient to describe the data. For all five baselined

data sets and the background-only corrected data this is what was observed. For the five baselined data sets the mean percentage X and Y variance captured in the PLS models ranged from 99.09 to 99.26 % and 99.80 to 99.85 % respectively. The difference in these values was not greater than the standard deviations reported and hence are not considered significant. Mean X and Y variance captured by the background-only corrected data set was 96.05 and 97.65 % respecively – less than that captured by the baselined data sets but still sufficiently described by one latent variable. Mean X and Y variance captured by the raw data was 99.03 and 54.67 %. This implies that the model could describe most of the variance in the spectral data but with only one latent variable, this variance was not sufficient to describe variance in concentration.

'Leave one out' cross validation statistics were reported for root mean square error (RMSE) and the correlation coefficient (R^2). A high correlation coefficient is desireable but does not necessarily reflect on how well a model is able to predict the parameter of interest. Root mean square error descibes how well the model fits the data and a low value is desirable. Reporting cross-validation statistics rather than calibration statistics gives some indication of the predictive capacity of the model. Again, only small differences were observed in the statistics for the baselined data: RMSECV and R^2 ranged from 0.0330 to 0.0360 and 0.9927 to 0.9914 respectively. The differences in these values were in the same order as the reported standard deviations. Background-only and raw data yielded noticeably poorer model statistics: RMSECV and R^2 were 0.1134 and 0.9573 respectively for background-only corrected data, and 0.3991 and 0.7945 respectively for the raw data.

From these statistics, it can be inferred that for this particular set of data, removing the known background contribution improved model performance. Subsequent baselining further improved the PLS models. No big difference was observed, however, according to which of the five trialled baselining methods was used.

5.3.4 Other Considerations

Other considerations to take into account when selecting a baselining method include the calculation time and the number and selection of parameters. Calculation times can vary greatly. For example, using the 'tic-toc' stopwatch function in MATLAB, the baselining calculation for 158 spectra took 2, 6, 59 and 525 seconds respectively for the ALS, APLS, SWiMA and AMM algorithms. Ideally, the calculations would be reasonably fast although

in fairness it should be noted that some of these techniques are relatively new and have not yet been optimised by the authors for speed.

The parameters for each baselining method investigated in this study were optimised on data set 1 by visual inspection (examples given in Appendix 2) before comparing the methods with each other. This introduces a high degree of subjectivity, however, so ideally the selection of parameters for a baselining method would be automated. In addition to removing operator bias, this would allow optimisation of parameters on a spectrum-by-spectrum basis rather than applying 'one size fits all' settings. SWiMA is the only method trialled here that was designed to be fully automated. In its current form, however, the automated stopping criteria were found not to work on our spectra and so the number of iterations needed to be specified. Fixing this problem and/or modifying the other techniques so that parameters can be optimised automatically would constitute significant improvements to these methods.

5.3.5 Summary and Conclusions

Our results indicate firstly that Raman spectra of EBPR bacteria that are corrected with both background and baseline removal perform better than background only-corrected or raw data with respect to appearance and qualitative and quantitative analysis. Of the variety of baselining methods trialled, RB, LM, Beier and RBE did not appear to fit a sample spectrum well whatever model parameters were employed. These techniques were thus not investigated in detail in this study. Of the remaining five baselining methods trialled, none seemed to outperform the others consistently or significantly in the qualitative and quantitative models applied to various data sets. Visually, the methods based on spectra-wide polynomial fitting (ModPoly and AMM) may show inconsistencies in how tightly they fit adjacent regions of the spectra due to the characteristics of this type of function. On the other hand, SWiMA and ALS appeared to fit adjacent regions of the spectra with similar tightness. SWiMA gave the least smooth baseline but (together with AMM) fit the spectra best in extreme regions. In contrast, the baselines calculated by ModPoly, ALS and APLS deviated from the spectra at the extremes in one or more cases. Since the fit at the extremes is important for the spectra investigated here, SWiMA was selected as the baselining technique to use for the remainder of this study.

In future work, for the reasons of superior smoothness and calculation time, ALS may be a preferable technique to use if the fit at the spectral extremes is either not so critical or could be improved. Another approach which may be worth investigating could be to use the

SWiMA algorithm to identify zero-signal regions in each spectrum and to connect these regions with a low-order polynomial. Finally, for full-automation of these preprocessing steps, methods for automatic removal of cosmic rays from spectra should be investigated and applied rather than relying on manual removal of these spectral artifacts.

Chapter 6

Raman Spectra of Individual EBPR Bacteria: A Qualitative Examination

6.1 INTRODUCTION 123

6.2 MATERIALS AND METHODS 123

6.2.1 Spectroscopic Measurements 123

6.2.2 Data Sets 123

6.2.3 Analysis 125

6.3 IDENTIFICATION OF KEY EBPR METABOLITES 126

6.3.1 Characteristic Spectra of Bacteria, Polyphosphate, PHAs and Glycogen 126

6.3.2 Variation in Peak Position of Key Metabolites between Samples 136

6.3.3 Identification of Key Metabolites with MCR 139

6.3.4 Morphology of Individual Cells 145

6.4 AUTOMATED IDENTIFICATION OF CANDIDATE PAOS 147

6.4.1 Identification of cPAOs Based on Polyphosphate Peak Height 148

6.4.2 Identification of cPAOs Based on PLS-DA 154

6.4.3 Comparison of Methods for Automated Identification of cPAOs 159

6.5 SUMMARY AND CONCLUSIONS 160

6.1 INTRODUCTION

The application of Raman spectroscopy to study EBPR bacteria is a recent development and to date, only a handful of studies have been published on this subject [68, 72, 74]. As discussed in Chapter 2, this technique has the potential to open up new avenues of research in EBPR. The existing studies have given little consideration to spectral analysis, however, and report simply on peak height data for selected Raman bands arising from metabolites present in the bacterial samples. In this study we conduct a qualitative examination of the Raman spectra of individual bacteria sampled from three different EBPR reactors. This examination uses both simple peak analysis methods and also applies, for the first time, multivariate tools to the analysis of Raman spectra of individual EBPR bacteria. The aims of the study are firstly, in light of the paucity of studies in this area, to confirm that Raman spectroscopy can be applied to identify the presence of key metabolites in individual EBPR bacteria. Building on this, the ways in which some multivariate tools – specifically multiplicative curve resolution (MCR) and partial least squares-discriminant analysis (PLS-DA) – might be employed to either improve analysis or extend the application of the technique in the context of EBPR are investigated.

6.2 MATERIALS AND METHODS

6.2.1 Spectroscopic Measurements

Bacterial samples were prepared for Raman spectroscopy, and Raman spectra were acquired according to the methods described in Chapter 4. Spectra were acquired with 200 second extended scans (data sets 1-3) or 90 second extended scans (data set 4). Raman spectra were preprocessed with the background removal and SWiMA baselining methods described in Chapter 5 before further analysis.

6.2.2 Data Sets

Four main datasets were used in this study and subject to qualitative examination by both univariate analysis based on peak height data, and multivariate data analysis using MCR and PLS-DA.

1. Calibration base set (n=86).

   The calibration set comprised 86 Raman spectra of EBPR bacteria sampled (in 2012) from two 4-Litre SBRs located in the research group of Professor Gu at the Department of Civil and Environmental Engineering, Northeastern University, MA. The spectra were selected on the basis that they covered all the 'metabolic states' (i.e. combinations of which metabolites were present) that were encountered in spectra acquired in this work. 15 spectra from each of the following metabolic states were selected: cells containing polyphosphate (PP), polyhydroxyalkanoate (PHA), glycogen (GLY), PP+PHA, and no metabolites (NONE). In addition, just 11 spectra were included that contained signals from PP+GLY since these were all that could be identified with this particular combination of metabolites. No spectra were identified that contained the combinations PHA+GLY or PP+PHA+GLY. In total, characteristic polyphosphate bands were evident in 41 of the spectra.

2. 'EDI100' (n=100)

   EDI100 comprised 100 Raman spectra of bacteria sampled from the two SBRs built and operated at the University of Edinburgh as part of this work. By inspection the number of spectra containing each metabolite were as follows: 9 'NONE', 82 PP, 7 PP+GLY, and 2 PP+PHA. In total, characteristic polyphosphate bands were evident in 91 of the spectra.

3. 'NE100' (n=100)

   NE100 comprised 100 Raman spectra of bacteria sampled (in 2013) from two SBRs located in the research group of Professor Gu at the Department of Civil and Environmental Engineering, Northeastern University, MA. By inspection the number of spectra containing each metabolite were as follows: 14 GLY, 10 'NONE', 8 PHA, 62 PP, and 6 PP+PHA. In total, characteristic polyphosphate bands were evident in 68 of the spectra.

4. 'SLO100' (n=100).

   SLO100 comprised 100 Raman spectra of EBPR bacteria sampled from the EBPR lanes at Slough Sewage Treatment Works (Thames Water, UK). By inspection the number of spectra containing each metabolite were as follows: 3 GLY, 16 'None', 72 PP, 9 PP+GLY. In total, characteristic polyphosphate bands were evident in 81 of the spectra.

6.2.3 Analysis

Univariate (based on peak height)

Univariate analysis of Raman spectra was used in the automatic identification of polyphosphate based on peak height data. This routine was performed in MATLAB using a custom-written script. In brief, the script carried out the following steps. First, the region 1760 to 1780 cm^{-1} was assumed to be signal free. The level of noise in each spectrum was defined by the range of intensities in this region. A conservative peak threshold was subsequently defined as three times the level of the noise. Next, two regions of interest encompassing the two major polyphosphate bands were defined and the spectra in these regions were smoothed with a Savitzky-Golay filter and then resampled at smaller intervals using a spline function. A linear baseline was constructed between the minima positioned either side of the peak maxima, and the peak height was defined as the peak maximum minus the baseline intensity. Peak position criteria were included so that if a peak position was found to be outwith the expected range of values, the output was set to zero. The peak position boundaries were calculated from the mean polyphosphate peak positions in the calibration base set plus and minus four times the standard deviation. A mutual exclusion criterion was also included so that if one polyphosphate peak was found to have an intensity set to zero (due to its position), the other peak intensity was also automatically set to zero.

Multivariate: Multiplicative Curve Resolution

The MCR model was built with PLS_Toolbox. Calibration data for the MCR model comprised the calibration base set described in Section 6.2.2 together with 8 spectra of pure PHB and 8 spectra of pure glycogen (making 102 spectra in total). Preprocessing comprised normalising each spectrum to its most intense peak so that each sample had similar influence on the model. The MCR model was built with 4 components (representing the four main constituents of the spectra – polyphosphate, glycogen, PHA, and bacterial component) that together accounted for 96.8% variability in the calibration spectra.

Multivariate: Partial Least Squares-Discriminant Analysis

The PLS_DA model was built with PLS_Toolbox. Calibration data for the MCR model comprised the calibration base set with each sample in the set assigned a candidate PAO (‘cPAO’) or no polyphosphate (‘noPP’) label based on the presence or absence of characteristic polyphosphate bands in each spectrum as identified by visual inspection. Preprocessing comprised normalising each spectrum to its most intense peak so that each

sample had similar influence on the model. The spectra were also mean centred. The model was cross-validated (venetian blinds, 10 splits). Based on the root mean square error of cross-validation (RMSECV) and variance captured in X and Y, the model was built with 3 latent variables.

6.3 IDENTIFICATION OF KEY EBPR METABOLITES

6.3.1 Characteristic Spectra of Bacteria, Polyphosphate, PHAs and Glycogen

Examples of Raman spectra that show the characteristic bands for bacterial spectra and three key EBPR metabolites – polyphosphate, PHAs and glycogen – are presented in the following sections. The position of the metabolite spectral bands originating from a pure compound and from within a bacterium are compared.

For quick reference, the structures of polyphosphate, two types of PHA (polyhydroxybutyrate (PHB) and polyhydroxyvalerate (PHV)), and glycogen are reviewed in Figure 6-1. Polyphosphate is a polymer comprising tens to hundreds of tetrahedral phosphate residues linked by high-energy phosphoanhydride bonds [147]. PHB is a polymer composed of 3-hydroxybutyrate units that in bacteria, are enzymatically formed by condensation of two acetyl-CoA units followed by a reduction step [148]. The residues are linked by ester groups and thus PHB is a type of polyester. PHV is identical to PHB except for the side chains which are methyl groups in PHB and ethyl groups in PHV. Whereas polyphosphate, PHB and PHV are linear polymers, glycogen is a highly branched polymer (polysaccharide) composed of glucose units linked by glycosidic bonds. At the core of the glycogen structure is an enzyme, glycogenin, which catalyses the first few polymerisation steps for each glycogen molecule.

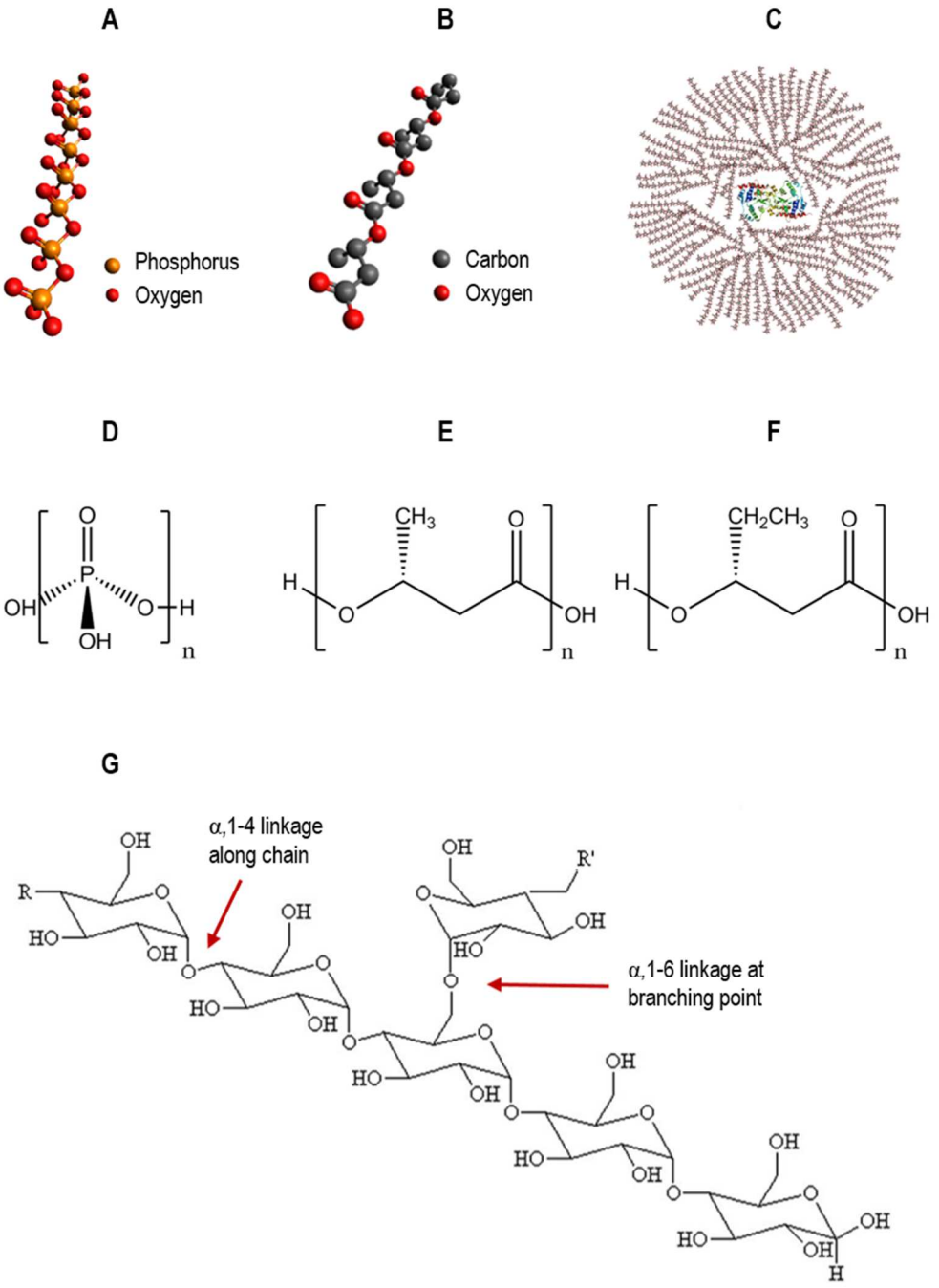


Figure 6-1 Pictorial representations and chemical structures of three key EBPR metabolites. Pictorial representations are of polyphosphate (A), PHB (B) and glycogen (C). Note that hydrogen atoms are not shown in A and B. Chemical structures are of polyphosphate (D), PHB (E), PHV (F) and glycogen (G). The image for glycogen is taken from [149].

Bacteria

Fundamentally, by (dry) weight, bacteria are composed of approximately 40-60 % proteins, 10-20 % carbohydrates, 10-15 % lipids and 7-19 % nucleic acids [150]. These constituents give rise to characteristic Raman bands [151]. Raman spectra of a dried cell-pellet smear of *Pseudomonas putida* (a gram-negative soil bacterium) and an individual EBPR bacterium containing no visible storage compounds are shown in Figure 6-2. Raman spectra of the two different bacterial species are shown to demonstrate the general spectral features common to all types of bacteria.

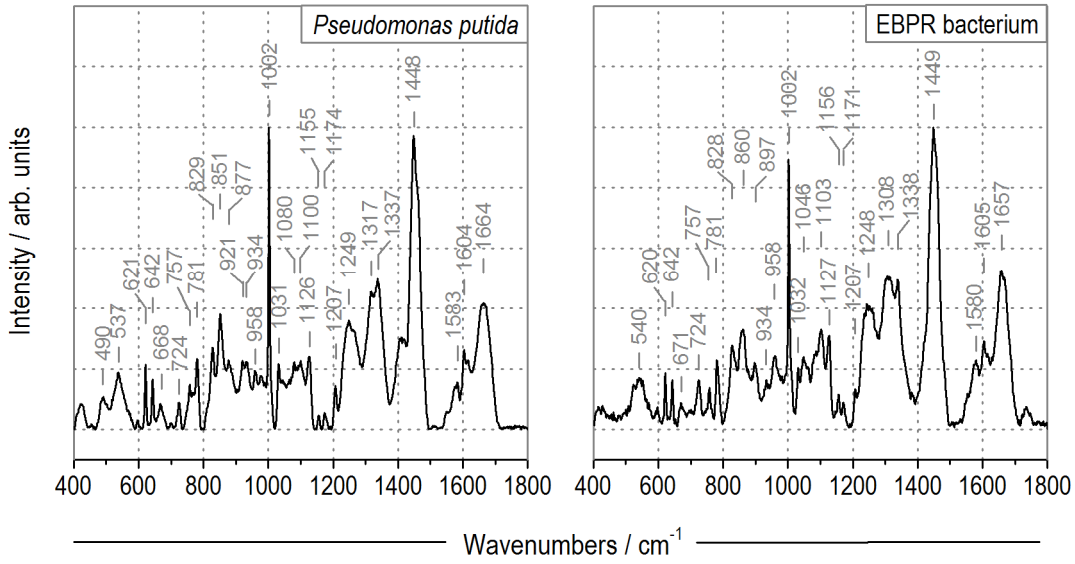


Figure 6-2 Raman spectra of a dried cell-pellet smear of Pseudomonas putida and of an individual EBPR bacterium. The positions of the major characteristic peaks are labelled.

Prominent features of bacterial Raman spectra include C-H stretching bands in the range 2700-3100 cm^{-1} (not shown) and a C-H deformation band at ~1450 cm^{-1}. These bands are commonly assigned to lipids, although it should be noted that CH, CH_2 and CH_3 functional groups are also present (albeit to a lesser extent) in the side chains of proteins and carbohydrates. Proteins are formed from chains of amino acids linked by –CONH– groups known as peptide (or amide) bonds. Two particular Raman bands assigned to these structures are commonly observed: the 'amide I band' at ~1650 cm-1 that mainly consists of a C=O stretching vibration, and the 'amide III band' occurring between 1250 and 1300 cm^{-1} that is a

mixture of C-N stretching and N-H bending [152]. Finally, a number of sharp bands understood to arise from the nucleotides guanine and adenine, and the amino acids tyrosine and phenylalanine can be observed at ~ 671, 724, 642 and 1002 cm^{-1} respectively. The phenylalanine band at ~1002 cm^{-1} is of particular note due to its high intensity. The positions and assignments of these peaks are summarised in Table 6-1. Although these are common Raman spectral features associated with bacteria, the exact position, intensity and shape of some of these bands may change according to factors such as bacterial species, growth conditions, cell cycle phase, and sample preparation.

Table 6-1 Position of major Raman spectrum peaks arising from a dried cell-pellet smear of Pseudomonas putida and an individual EBPR bacterium.

Wavenumber / cm^{-1}		Assignment	Reference
P. putida	**EBPR bacterium**		
621	620	C-C twisting mode of phenylalanine	[153]
642	642	C-C twisting mode of tyrosine	[153]
668	671	guanine	[150]
724	724	adenine	[150]
1002	1002	C-C aromatic ring stretching phenylalanine	[153]
1249	1248	amide III (1250-1300 cm^{-1})	[152]
1448	1449	C-H_2 deformation (1440-1460 cm^{-1})	[150]
1604	1605	C=C stretch (protein)	[153]
1664	1657	amide I (1650-1680 cm^{-1})	[150]

Polyphosphate

Raman spectra of sodium phosphate glass type 45 (SPG45) and polyphosphate in an EBPR bacterium are shown in Figure 6-3. Acquiring a well-characterised pure compound of polyphosphate that is analogous to that found in bacteria is challenging. Here we used SPG45 – a polyphosphate supplied by Sigma-Aldrich (product number S4379) with the chemical formula $Na_{n+2}P_nO_{3n+1}$, where n = 45, the average number of phosphate residues in the chain. Since completing this work, this product has been discontinued along with sodium phosphate glass types 5, 15, 65 and 75+ (product numbers S5878, S6003, S6253 and S8262 respectively) which have been used in previous studies [154, 155].

The most intense Raman peak in the SPG45 spectrum is observed at around 1159 cm^{-1} and this arises from symmetrical stretching of the PO_2^- species [156]. These moieties are not

included in the polymer backbone (Figure 6-4) leaving them susceptible to hydrogen bonding in aqueous solution and a resultant peak shift to lower wavenumbers. For short-chain polyphosphates, an increase in chain length was seen to shift the PO_2^- peak to higher wavenumbers – from 1022 cm^{-1} for pyrophosphate to 1157 cm^{-1} for hexametaphosphate [68].

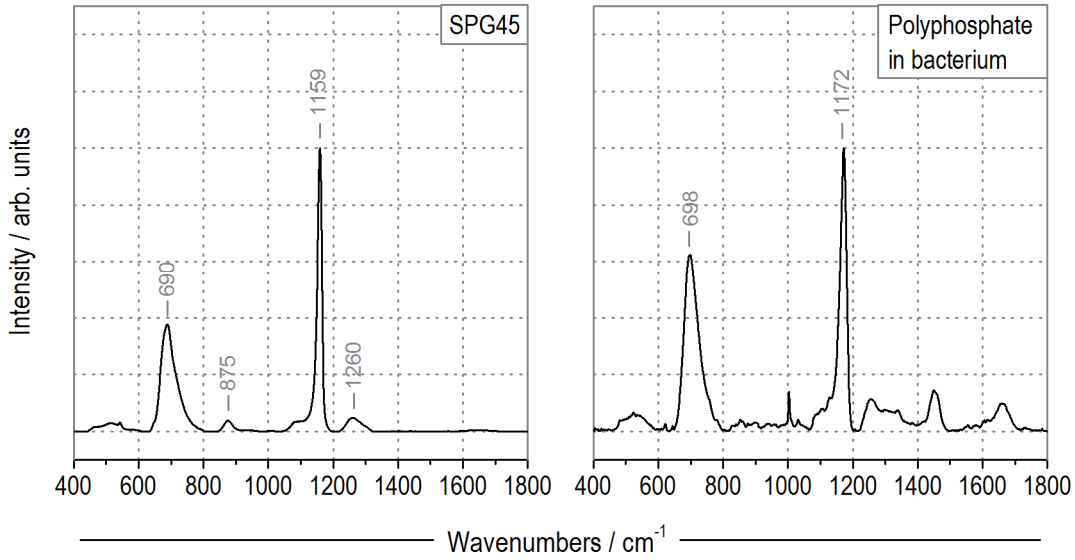


Figure 6-3 Raman spectra of sodium phosphate glass type 45 (SPG45) and polyphosphate in an EBPR bacterium. The positions of the major characteristic peaks of polyphosphate are labelled.

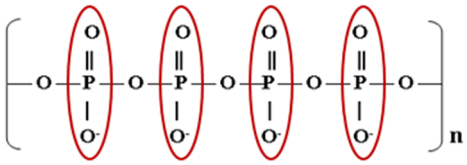


Figure 6-4 Linear structure of polyphosphate with PO_2^- moieties circled in red.

The SPG45 Raman peak at ~690 cm^{-1} arises from vibrations of the P-O-P group which form the polymer backbone [157]. This peak is broad due to the amorphous nature of the polymer, with the atoms around the backbone adopting many different orientations. The P-O-P peak position moves to lower wavenumbers with increasing chain length of the polymer species [157] and this was shown by Majed *et.al.* for short chain polyphosphates (≤ 6 phosphate units) [68]. The major peak positions of SPG45 reported here (690 and 1159 cm^{-1}) are very

similar to those reported by Majed *et al.* for hexametaphosphate (690 and 1157 cm^{-1}), however, suggesting that shifts in peak position due to chain length could be very small or negligible with longer-chain polyphosphates. Furthermore, a polymer is a mixture of molecules with different chain lengths so that changes in the P-O-P peak position due to changes in chain length may be impossible to detect.

The spectrum of polyphosphate in an EBPR bacterium shows that both major polyphosphate peaks are clearly visible and are approximately ten wavenumbers higher than the same bands in SPG45. This means that the P-O-P and PO_2^- vibrations are higher in energy than the same vibrations in SPG45. The relative positions of the peaks are summarised in Table 6-2. If this change were solely due to longer-chain polyphosphates being present in bacteria as suggested by Majed *et al*., we might expect (based on the previous studies discussed) the positions of the peaks to be shifted in opposite directions: P-O-P to a lower wavenumber and PO_2^- to a higher wavenumber. Since both peaks are shifted in the same direction, however, it suggests the underlying reason for this is more complex. Very little is known about the biochemistry of polyphosphate in EBPR biomass where it may be associated with intracellular membranes or complexed with DNA, RNA or proteins [11]. Any of these associations or complexes can have an impact on the distribution of electrons in polyphosphate bonds, and hence the bond strengths and ultimately the Raman spectra.

Table 6-2 Position of major Raman spectrum peaks arising from a pure compound (SPG45) and from polyphosphate within an EBPR bacterium.

Wavenumber / cm^{-1}		Assignment	Reference
SPG45	Polyphosphate in bacterium		
690	698	P-O-P symmetric stretch	[156, 157]
1159	1172	PO_2^- symmetric stretch	[156, 157]

Polyhydroxyalkanoates

Raman spectra of PHA in an EBPR bacterium and of the pure compounds PHB and PHB-co-(12%)-PHV (both are types of PHA) are shown in Figure 6-5. These two pure compounds were measured since different metabolic pathways within EBPR bacteria can potentially build PHAs from four different monomeric units – 3-hydroxybutyrate, 3-hydroxyvalerate, 3-

hydroxyl-2-methylbutyrate, and 3-hydroxyl-2-methylvalerate [26]. When the only carbon source available during EBPR is acetate, the intracellular PHAs are built predominantly from 3-hydroxybutyrate, and hence PHB is the main polymer synthesised.

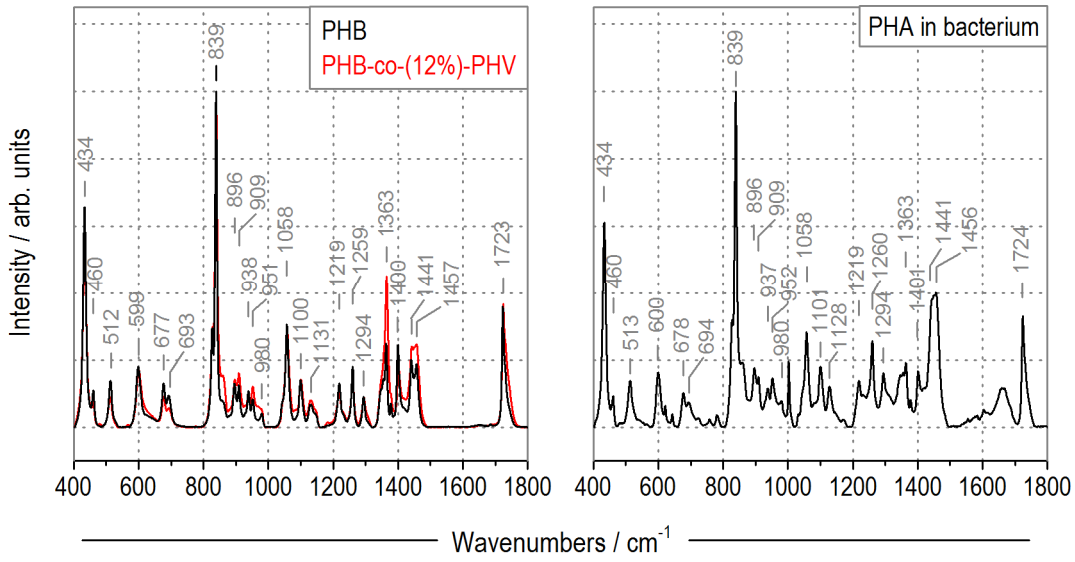


Figure 6-5 Raman spectra of the pure compounds polyhydroxybutyrate (PHB) and polyhydroxybutyrate-co-(12%)-polyhydroxyvalerate (PHB-co-(12%)-PHV) and of an EBPR bacterium containing PHA. The positions of the major characteristic peaks of the PHAs are labelled.

It can be seen that the Raman spectra of pure PHAs exhibit many peaks distributed throughout the wavenumber range shown here, although it is noticeable that the region between 1500 and 1700 cm^{-1} is somewhat 'silent'. The three strongest peaks in the pure compounds are narrow with respect to polyphosphate bands, and are positioned at ~434, ~839 and ~1723 cm^{-1}. These have been assigned to skeletal C-C deformations, C-COO stretching, and C=O stretching respectively [65, 97]. In the EBPR bacterium spectrum these three strongest PHA bands are clearly visible and the peak positions are not shifted with respect to the pure spectra. The assignments and relative positions of the PHA peaks in a pure compound versus the bacterium are summarised in Table 6-3.

Table 6-3 Position of major Raman spectrum peaks arising from pure PHB and from PHA within an EBPR bacterium.

Wavenumber / cm^{-1}		Assignment	Reference
PHB	PHA in bacterium		
434	434	(CC) skeletal deformation	[65]
677	678	(C=O) out-of-plane bending	[97]
693	694	(C=O) out-of-plane bending	[97]
839	839	(C-COO) stretching	[65, 97]
1058	1058		
1100	1101	(CC) skeletal stretch	[65]
1131	1128		
1219	1219		
1259	1260		
1294	1294	(CH), (CH2), (CH3)	[65]
1363	1363	deformations	
1400	1401		
1441	-		
1457	1456		
1723	1724	(C=O) stretch	[65]

The literature on application of Raman spectroscopy to study PHAs in bacteria is growing due to industrial interest in marketing PHAs as biodegradable plastics [158]. De Gelder *et al.* [65] proposed that the Raman band intensity at ~1734 cm^{-1} was suitable for monitoring PHB production and consumption in *Cupriavidus necator* DSM 428 (H16) due to its isolation from other bands. Izumi and Temperini [97] used FT-Raman to investigate the differences in the Raman spectra of PHB and PHB-co-PHV. They demonstrated that the relative intensity of bands at 1443 and 1458 cm^{-1} as well as the FWHM of the band at 1725 cm^{-1} were indicators of the crystallinity of the material rather than the composition. For molten polymers, however, the ratio of bands at 1354 and 1740 cm^{-1} could be used to estimate the composition of PHB-co-PHV samples i.e. the percentage of PHV co-monomer present.

Comparing the Raman spectra presented here of PHB and PHB-co-(12%)-PHV, the described changes in bands at ~1441, ~1457 and ~1723 indicating changes in crystallinity can be seen. In addition there is a marked difference in the ratio of the bands at ~1363 and 1723 cm^{-1} which are perhaps analogous to the band ratios suggested by Izumi and Temperini for monitoring composition change in molten polymers. To be able to distinguish between PHB and PHV in EBPR bacteria using Raman spectroscopy would extend the investigatory power of this technique. The practical feasibility of this needs to be investigated, however. In

the PHB-in-bacteria spectrum shown here, the band at ~1363 cm^{-1} is in a region of the Raman spectrum that encompasses broad background signal contribution from the underlying bacterial components so a ratiometric peak height method may not be reliable for determining the type of PHAs present in bacteria. The use of multivariate analysis may be able to overcome this.

One way to investigate this may be to feed the EBPR bacteria with propionate and compare the signature PHAs present in the Raman spectra with those acquired from bacteria fed with acetate. This was attempted in a short experimental test – where the bacteria in one SBR were fed with propionate for one cycle and samples were taken at the end of the anaerobic phase for analysis. Unfortunately, in 100 Raman spectra acquired, however, none were observed to contain any type of PHA and so the comparison could not be made at the present time.

The reason that no PHA was observed in the test described above is not yet understood. Previous studies have shown that PAOs are capable of adapting their metabolism quickly to be able to assimilate acetate or propionate, so much so that 'switching' the feed composition in this manner has been recommended as a strategy for giving PAOs a competitive advantage over GAOs [159, 160]. Furthermore, in this study the increase in phosphate concentration and decrease in COD concentration observed in the mixed liquor during the anaerobic phase indicated that in the same way as acetate, the supplied propionate was assimilated by the biomass. An assumption was made, therefore, that the assimilated propionate was formed into PHA for storage since it could not be respired by PAOs in the absence of oxygen or nitrate. Perhaps this assumption was false. Alternatively, it could be the case that PHA was present in some bacteria but was below the (as yet undetermined) limit of detection for this method. Another explanation could be that none of the bacteria examined contained PHA and that experimental bias for nicely presented cells on the slide could have contributed to this particular sample selection and outcome. Since the experiment was attempted only once, however, it is intended that it will be repeated again in future work and depending on the outcome, these points may be investigated further.

Glycogen

Raman spectra of pure glycogen and glycogen in an EBPR bacterium are shown in Figure 6-6. Glycogen is a large, amorphous polymer giving rise to several broadened Raman bands. The strongest of these is at ~480 cm^{-1} and is within the range 280 to 515 cm^{-1} that was assigned to "skeletal modes" by Galat [161]. This is a relatively low energy region of the Raman spectrum so the vibrational modes giving rise to peaks are likely to be complex rather than arising from any easily distinguishable bond or functional group.

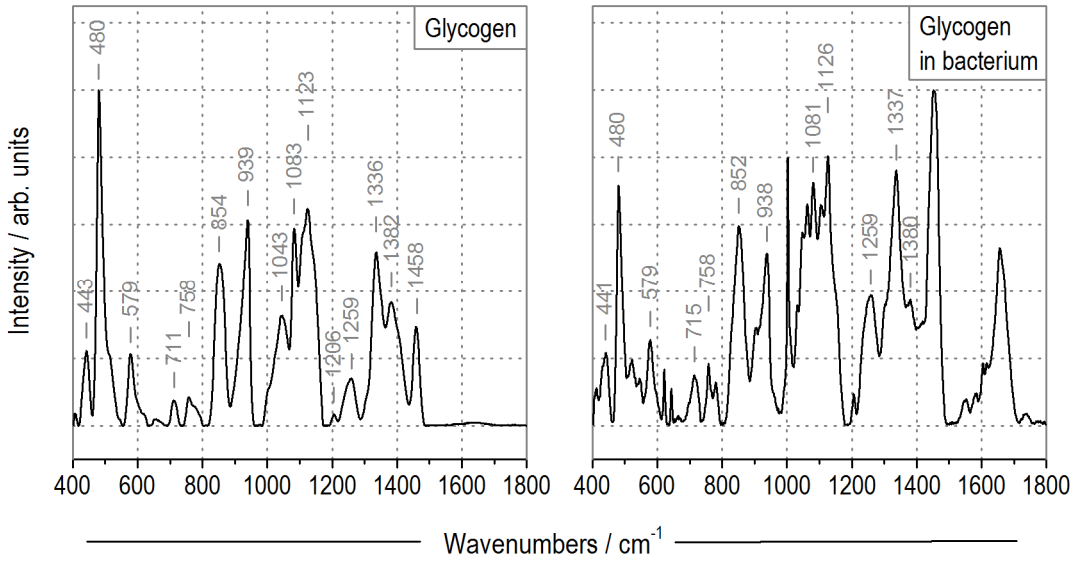


Figure 6-6 Raman spectra of pure glycogen and glycogen in an EBPR bacterium. The positions of the characteristic glycogen Raman peaks are labelled.

It can be seen that there is no easiliy distinguishable difference between the peak postions of Raman signals arising from pure glycogen and from glycogen in the EBPR bacterium. The relative positions of the peaks are summarised in Table 6-4. Several Raman bands, particularly those at ~480, ~852 and ~938 cm^{-1} are easily recognisable in the bacterial spectrum. Other glycogen peaks, like those between 1000 and 1500 cm^{-1} are evidently superimposed on Raman bands from the bacterial components. Konorov *et al.* recently used the glycogen peak at 480 cm^{-1} to develop a Raman spectroscopy absolute quantification method for glycogen content in human embryonic stem cells [162].

Table 6-4 Position of major Raman spectrum peaks arising from pure glycogen and from glycogen within an EBPR bacterium.

Wavenumber / cm^{-1}		Assignment	Reference
Glycogen	Glycogen in bacterium		
480	480	skeletal modes	[161]
854	852	C(1)-H(α) bending modes	[161]
939	938	ring modes	[161]
1083	1081	COH bending, CH bending	[163]
1123	1126	C-C stretching	[163]
1336	1337	CH bending	[163]
1382	1380	CH bending	[161]
1458	-	CH_2 bending	[161]

6.3.2 Variation in Peak Position of Key Metabolites between Samples

For each key metabolite discussed (polyphosphate, PHA, glycogen), the position of the most intense peaks in the Raman spectra were compared across all the samples analysed in this chapter i.e. in all samples from all four data sets. The results are summarised in Table 6-5.

Table 6-5 Variance in the position of the most intense Raman bands for each key metabolite across all bacterial samples analysed in this Chapter 6.

Peak Description	Number of Samples	Mean Position / cm^{-1}	Standard Deviation / cm^{-1}	Position in Pure Compound
Phenylalanine	372	1002.5	0.6	-
Glycogen	58	481.1	1.0	480
Polyphosphate 1	277	696.3	2.0	690
Polyphosphate 2	277	1168.5	3.5	1159
PHA 1	46	433.2	0.5	434
PHA 2	46	838.8	0.8	839
PHA 3	46	1724.1	1.6	1724

For comparison, the position of the most intense peak commonly assigned to phenylalanine was also included in this analysis. The mean position of this across 372 samples was found to be 1002.5 cm^{-1} with a standard deviation of 0.6 cm^{-1}, indicating a very small variance between samples relative to the other peaks analysed. In Chapter 4, the standard deviation in the position of a silicon peak at 520.2 cm^{-1} measured during instrument set-up was found to vary with a standard deviation of 0.1 cm^{-1} (across 64 samples measured on different days). The silicon sample measured each time was the same well-controlled (crystallographic orientation), inorganic sample. The cell samples studied here, however, are poorly-controlled biological samples. The standard deviation of 0.6 cm^{-1} observed in the position of the phenylalanine peak can perhaps, therefore, be considered to approach the experimental error in measurement.

The number of samples containing glycogen and PHA were 58 and 46 respectively. Despite this lower number of samples with respect to phenylalanine (372 samples), the standard deviation observed in major peak positions for these metabolites was less than 1.0 (i.e. also low) with the exception of the PHA band at 1724.1 cm^{-1} which had a standard deviation of 1.6. Furthermore, the major bands for pure glycogen and pure PHB as measured earlier in this chapter are at the same position of the mean position measured in all samples within 2 standard deviations. These observations confirm the earlier observation that the characteristic Raman bands for glycogen and PHA change little between pure compounds and bacterial samples i.e. they show negligible matrix effect. Additionally, there is little variation observed in peak position between samples.

The position of the major polyphosphate Raman bands are shown to vary noticeably between samples. The mean positions of polyphosphate bands 1 and 2 were found to be 696.3 cm^{-1} and 1168.5 cm^{-1} with standard deviations of 2.0 and 3.5 respectively. This is much higher than the variance observed for peak positions of glycogen and PHA despite there being a much greater number of samples (277) analysed. In addition, mean peak position of both polyphosphate bands are not found to be the same (within 2 standard deviations) as those measured in the pure compound (SPG45). These observations confirm the earlier observation that there is a noticeable matrix effect for polyphosphate between the pure samples and bacterial samples. Furthermore, there is significant variation in peak position between samples with respect to that observed for glycogen, PHA and phenylalanine.

A ‘one way between subjects’ analysis of variance (ANOVA) test was applied to the samples to determine if the mean polyphosphate peak positions were significantly different

at the $p<0.05$ level depending on the sample origin or the particular combination of metabolites present in each sample (PP only, PP+PHA or PP+GLY).

Within the calibration data, a significant difference in the position of both polyphosphate peaks was found according to which combination of metabolites were present. An overview of the peak position statistics can be seen in the box plots in Figure 6-7. It should be noted, however, that the sample numbers were small: 15, 15 and 11 respectively for the PP only, PP+PHA and PP+GLY groups. Still, it would be interesting to investigate this further in future samples.

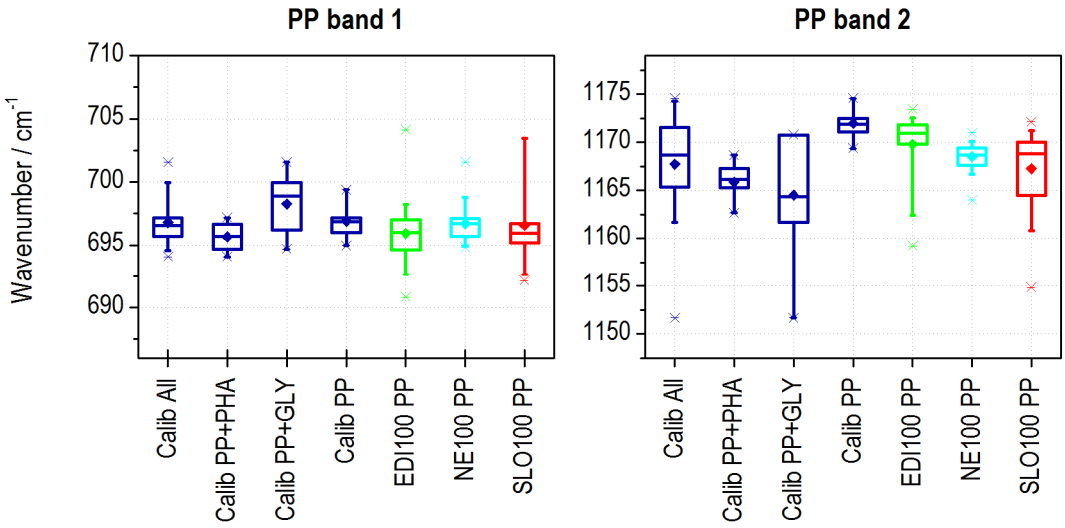


Figure 6-7 Box plots showing statistics for peak position of polyphosphate bands 1 (at ~697 cm^{-1}) and 2 (at ~1169 cm^{-1}) in the subsets of bacteria indicated. Each box encapsulates the 25th to 75th percentile and the whiskers show the limits of the 5th to 95th percentile. The line in the centre of the box represents the median value and the point in the box represents the mean value. The stars indicate outliers.

Between the four data sets (Calibration, EDI100, NE100 and SLO100), no significant difference was found in the mean position of the polyphosphate peak at ~696 cm^{-1} for samples containing only PP. For the peak at ~1169 cm^{-1} however, the difference in mean peak positions for the four groups was found to be significant. For these four groups, the samples sizes were 15, 82, 62 and 72 respectively. These data statistics can also be seen in Figure 6-7. It can be concluded reasonably that the polyphosphate peak at ~1169 cm^{-1} is

more sensitive to matrix effects than the peak at ~696 cm^{-1}. A much larger sample set would be required, however, to be able to conclude with confidence that the sample source can influence the biochemistry (and hence 'matrix' of an EBPR bacterium as to influence the Raman band position of polyphosphate in that bacterium.

Overall, the variance in polyphosphate peak position could raise some challenges for the analysis of EBPR bacteria by Raman spectroscopy. For example, if quantitative analysis of polyphosphate content were desirable, what kind of compound could be used as a standard when a pure polyphosphate sample has such a different signal to that arising from intracellular bacterial polyphosphate? On the other hand, perhaps these peak position shifts could be a useful indicator of some parameter in the cells themselves. For example, in the bacteria studied here the mean position of the two polyphosphate peaks were 696 and 1169 cm^{-1}. The polyphosphate Raman signals arising from yeast vacuoles (of *Candida albicans*), however, were reported to be at 688 and 1154 cm^{-1} [70]. These peak positions are more similar to those measured in pure SPG45 in this work (690 and 1159 cm^{-1}). This variance in polyphosphate peak positions is something that could be considered in future work.

6.3.3 Identification of Key Metabolites with MCR

The spectra of EBPR bacteria are complex with many overlapping signals. Despite this, multivariate statistical tools have not yet been applied to these kind of samples. For example, multiplicative curve resolution is designed to extract underlying pure component spectra and related concentrations from a data set with as few assumptions about the data as possible. The algorithm is not limited to finding orthogonal components (like PCA), but is limited to producing components that are positive. These conditions mean that the model components are more likely than PCA to be physically meaningful in that they represent real chemical components contributing to the overall spectrum rather than abstract combinations of variables.

The results of applying MCR analysis to our data are shown below. The model was built on the base calibration data set described in Section 6.2.2. In addition, 16 spectra of pure compounds (8 PHB and 8 glycogen) were added to the calibration set to aid (bias) the MCR model in finding physically meaningful solutions. The spectrum of pure polyphosphate was not included in the calibration data set since this spectrum, as discussed, is significantly different from the polyphosphate Raman bands found in the sample spectra. The model was

built with 4 components, selected to represent the four main constituents of EBPR bacteria: polyphosphate, PHA, glycogen and the underlying bacterial matrix.

The 4 components resulting from MCR analysis are shown in Figure 6-8 and together these captured over 95% of the variance in the Raman spectra. It can be seen that the model was successful in producing components that closely resemble the pure spectra of the three key metabolites – polyphosphate, PHA and glycogen (shown in Section 6.3.1) – plus a spectrum that could reasonably be considered to represent the underlying bacterial 'matrix'. One notable difference is that the PP-type component also contains small amounts of characteristic bands from bacterial components such as phenylalanine (~1002 cm^{-1}), C-H_2 deformation (~1450 cm^{-1}), and amide I (~1657 cm^{-1}). Relative to this, the PHA- and GLY-type components contain very little visible contribution from these kind of bacterial signals. This is partly due to the inclusion of pure compound spectra of both these metabolites in the calibration data set.

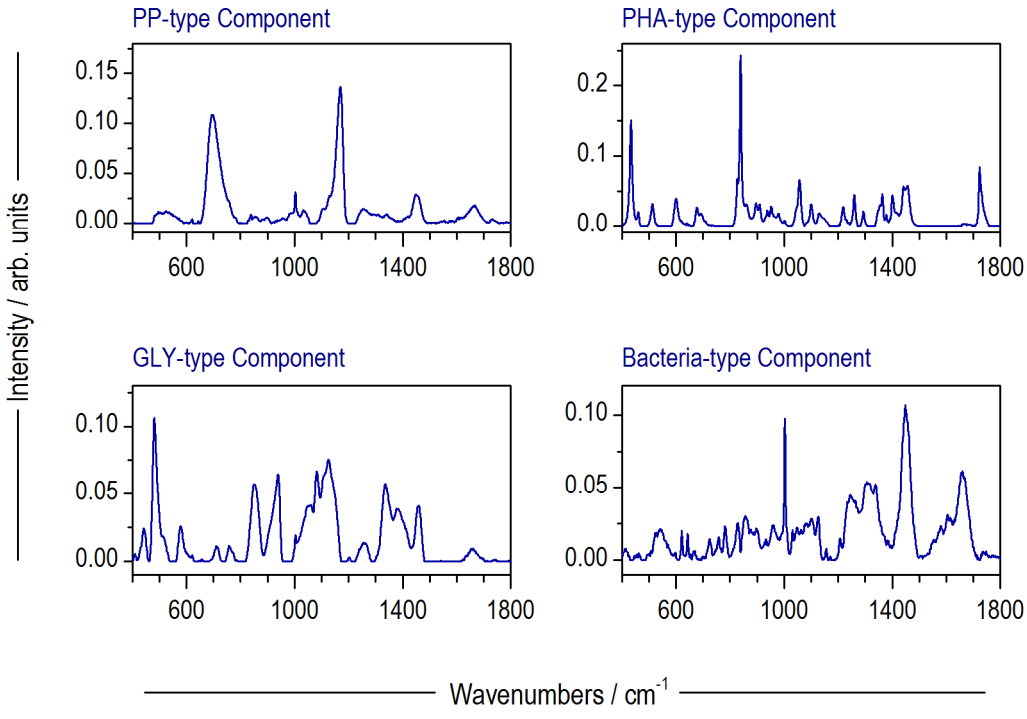


Figure 6-8 Loadings for the 4-component MCR model built as described in Section 6.2.3.

The MCR model was applied to three further datasets; EDI100, NE100 and SLO100. Scores for each sample (Raman spectrum) from each dataset on each of the four loadings are shown in Figure 6-9. The scores plots are arranged to show three different combinations of loadings: PHA-PP, GLY-PP and GLY-PHA. During preprocessing the spectra were normalised by setting the most intense peak in each spectrum to a value of 1. This means that the scores are only directly comparable when the peak with the largest intensity is the same between samples. The scores still give, however, a reasonably clear visual representation of which of the three key metabolites are present in each sample.

Generally, the samples are positioned where expected on the plots according to which metabolites were observed in each spectrum by visual inspection. One notable exception to this are the data points for glycogen-containing spectra which are plotted in a position such as to suggest the presence of low levels of PHA in the sample spectra. It was known, however, that no samples contained both PHA and glycogen in these datasets. An approximate threshold to suggest the presence of PHAs would thus likely have to be set higher than that for PP or GLY.

It should be noted that although the combinations of PHA+GLY and PP+PHA+GLY were not observed in any of the bacteria examined in this work, they have been observed in previous work [72, 74]. These intracellular compounds can, therefore, be present simultaneously. During the anaerobic phase of an EBPR cycle, intracellular concentration of PHA increases whilst GLY decreases. Subsequently, during an aerobic phase, GLY increases whilst PHA decreases. It is likely that none of the bacteria analysed in this work were in the ‘crossover point’ where both metabolites were present simultaneously – since they were sampled from either the end of the anaerobic phase or the end of the aerobic phase. Furthermore, it could be possible that the concentration of these compounds during the crossover point is below the limit of detection for our instrumentation. These are matters for further investigation.

The sample scores for the calibration dataset are evenly spread and cover the plot space (i.e. the variable space) well. This is not surprising since these are the data on which the MCR model was built. By contrast the scores plots for the other three data sets are less distributed over the variable space. The EDI100 dataset for example, is limited to just four sample types (PP, PP+GLY, PP+PHA and NONE), and of these the majority (82 samples) contain just PP. This gives the plots a very sparse appearance in comparison to the calibration dataset.

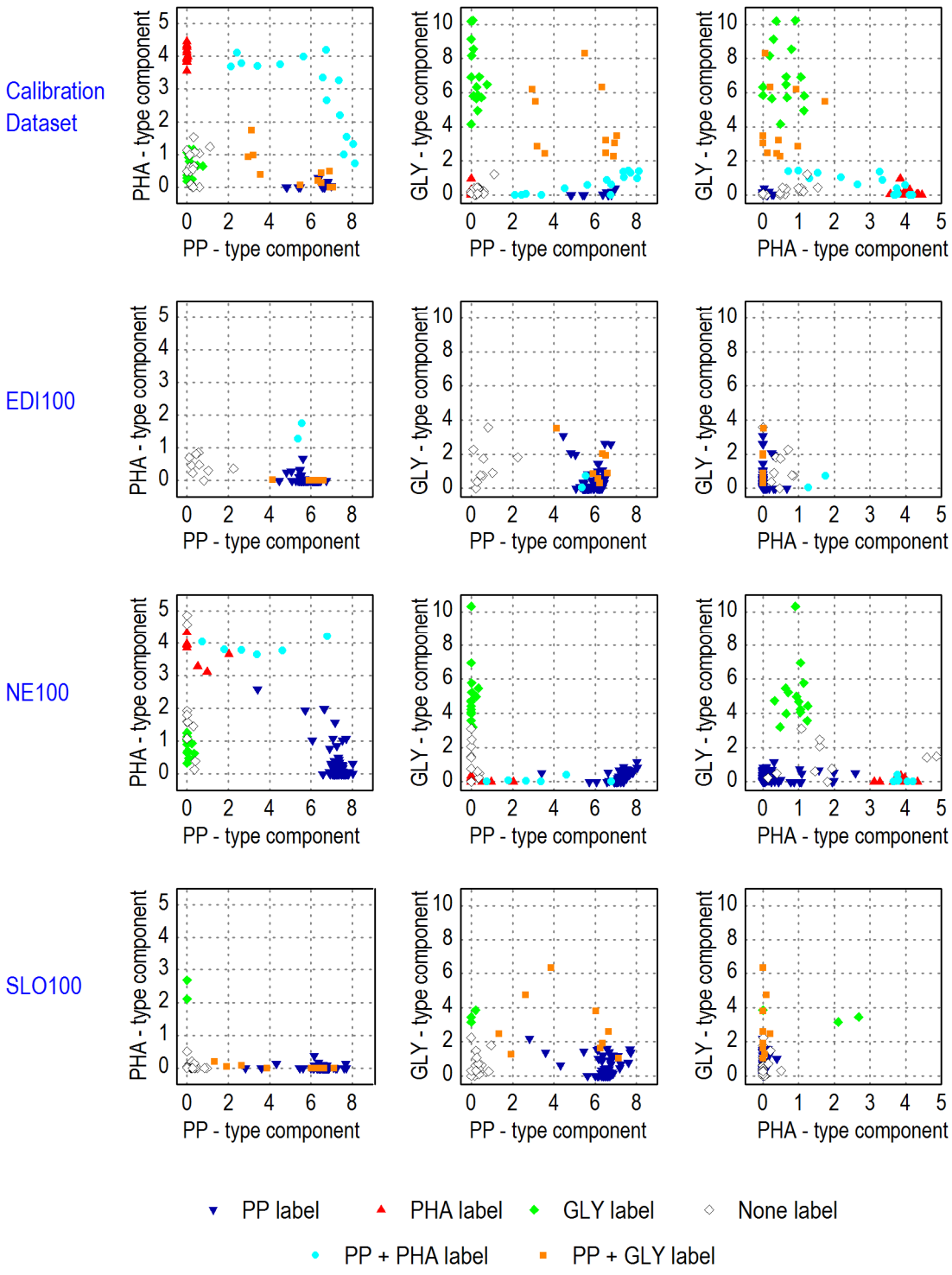


Figure 6-9 Scores plotted for a 4-component MCR model built as described in Section 6.2.3 on a calibration dataset (top), and applied subsequently to the EDI100, NE100 and SLO100 datasets. Each sample is represented by a symbol which is coloured according to which metabolites were present in the spectra by visible inspection.

Overall it can be seen that this kind of analysis is perhaps not exact enough for definitively categorising which metabolites are present in each sample spectrum analysed. The scores plots do, however, give an overall qualitative impression of the combination of metabolites present in each sample and the distribution of these among a data set. For the analysis of individual EBPR bacteria, it is likely that data sets comprise large numbers of spectra so that meaningful comparisons can be made. For this scenario, this kind of analysis may prove useful for quick comparison between datasets. Furthermore, it would be interesting to compare these plots for samples analysed under different operating conditions. Perhaps they would constitute a kind of 'fingerprint' that is characteristic of a particular reactor or operating conditions and may then further be related to process performance. Finally, the success shown by MCR in extracting components so similar to the pure compound spectra and the underlying bacterial matrix is something that may prove useful for quantitative analysis of these spectra. This is investigated in Chapter 7.

One advantage of using multivariate statistics is the possibility of residuals analysis – reviewing spectral contributions that could not be fit by linear combinations of the model loadings. For the MCR model described above, the summed original spectra for each data set are shown in Figure 6-10, overlaid with the summed spectra that were reconstructed by the model. The difference between the original and reconstructed spectra i.e. the summed residuals are also shown.

The residuals for the calibration data set are noticeably less in magnitude than those for the other three data sets. This shows that the model, not surprisingly, fits the calibration data set on which it was built better than the other data sets to which it was applied. For ease of comparison of peak position, the summed residuals for all four data sets are shown overlaid in Figure 6-11. Some peaks in the residuals, such as those at ~1180 cm^{-1} may be present due to slight difference in relative peak position for a metabolite, in this case one of the polyphosphate peaks. Other peaks in the residuals, such as those at ~970 cm^{-1} (EDI100), ~1050 cm^{-1} (NE100, SLO100), ~1090 cm^{-1} (EDI100, SLO100) and ~1380 cm^{-1} (EDI100), however, are more distinct from the known strong bands arising from EBPR metabolites bands, and if consistently present in different bacteria may be worth investigating further. At the present time, however, the origin of these peaks is unknown.

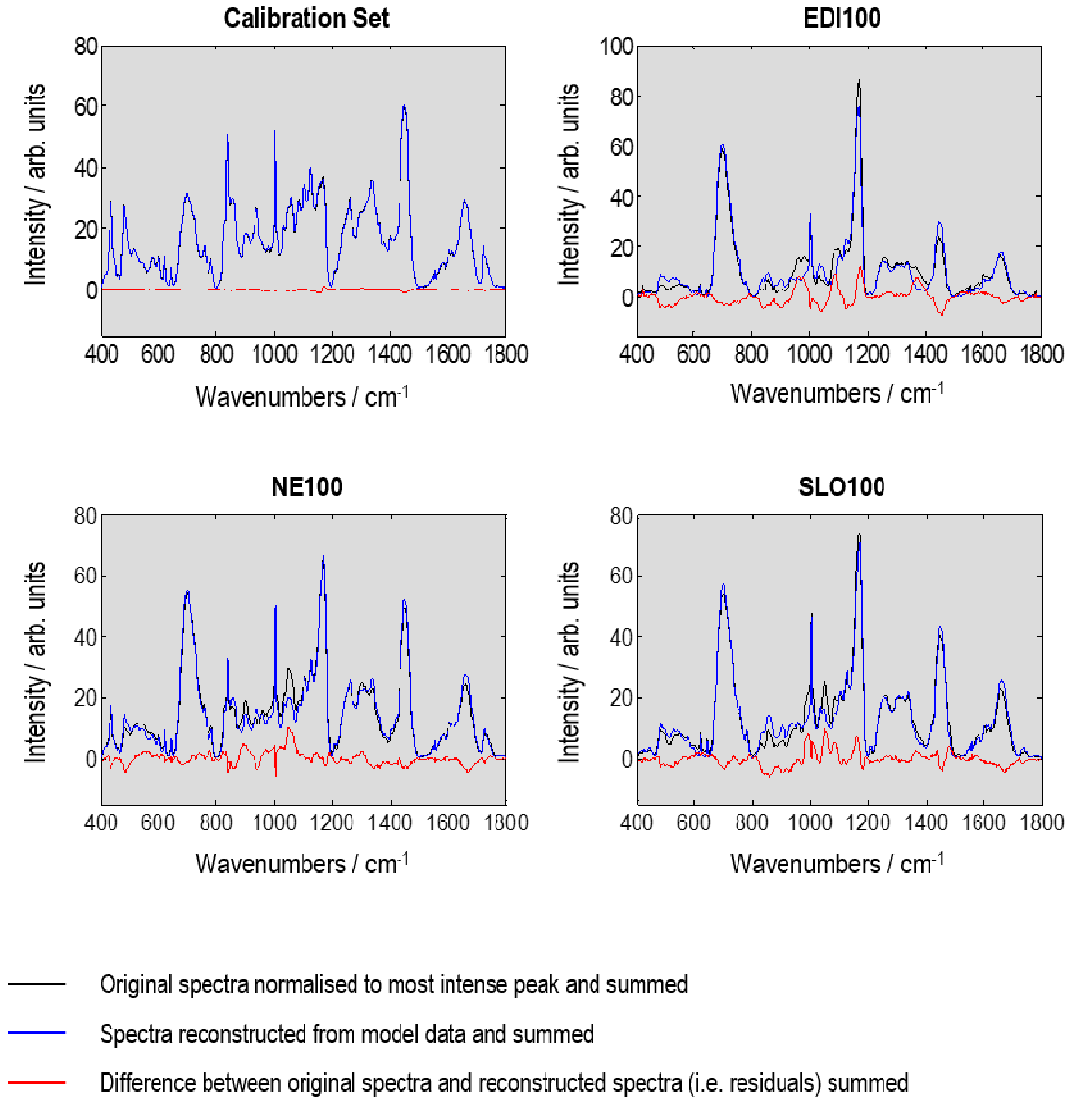


Figure 6-10 Comparison for each data set of original (summed) spectra and the (summed) spectra reconstructed from the MCR model. The difference between these two i.e. the summed residuals, are shown in red.

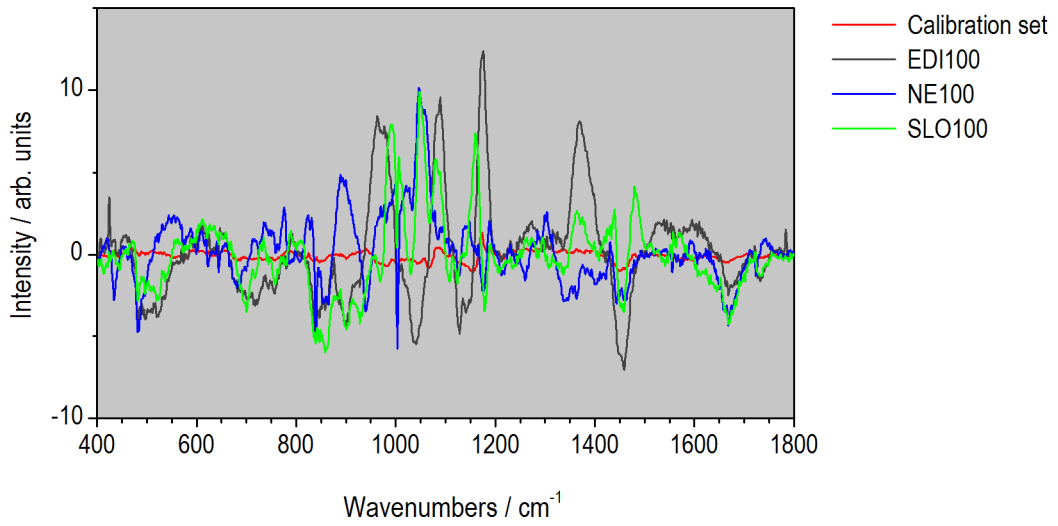


Figure 6-11 Summed residuals from the MCR model applied to each of the four data sets

The presence of residuals can be a help and a hindrance in data analysis. A residual contribution common to many samples can indicate an important component in the samples that has been overlooked and not represented sufficiently in the data model. Samples with large, significant residual contributions ('outliers') can indicate unusual samples that require closer inspection. These kind of samples, however, can cause errors in the model fit to that particular sample and therefore should be identified and removed from a dataset prior to analysis.

6.3.4 Morphology of Individual Cells

The morphology of individual EBPR bacteria has rarely been discussed in detail, and never with respect to the metabolites present in each bacterium. For this reason, example micrographs of individual EBPR bacteria analysed with Raman spectroscopy are shown in Figure 6-12. Where more than one cell is present, it is the cell in the centre of the image that was analysed with Raman spectroscopy. The micrographs are arranged according to sample source (i.e. dataset) and to which metabolites were visible in the bacteria. The examples were selected randomly and are representative of the typical appearance of bacteria analysed in this work. Three distinct morphologies can be seen in the samples: individual cocci, diplococci-type shapes where cells appear to be dividing, and tetrads. The cells shown range in size between approximately 1 and 4 µm in their smallest dimension.

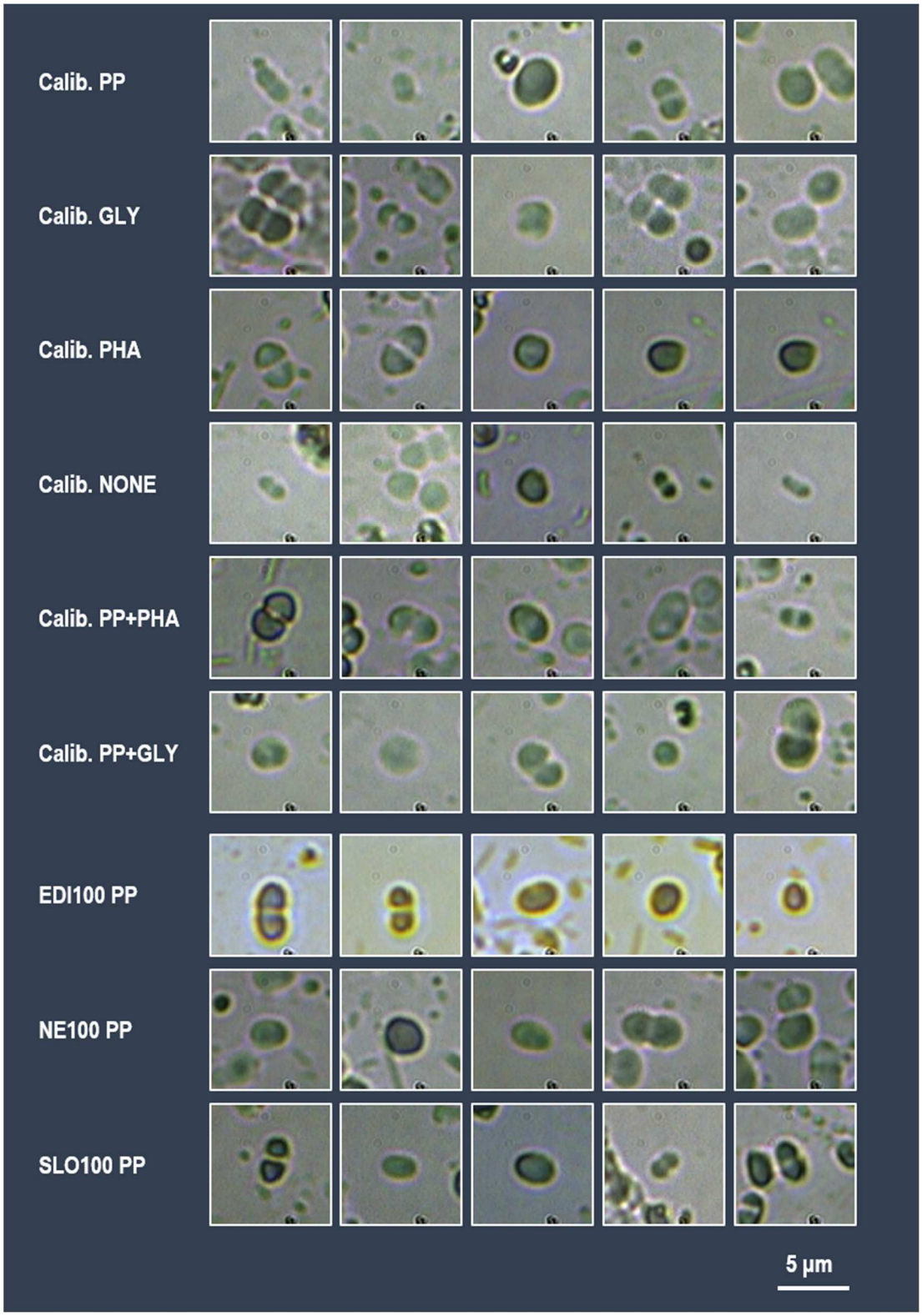


Figure 6-12 Example micrographs (magnification ×1000) of individual EBPR bacteria dried on a CaF_2 substrate and subsequently analysed with Raman spectroscopy. The micrographs are arranged according to sample source and which metabolites were visible in their Raman spectra.

PAOs have previously been described as having coccus or rod-shaped morphology [40, 164]. Samples identified as containing polyphosphate (and hence candidate PAOs) in this work were found to have either cocci or dividing-cocci morphologies. No rod-shaped cells containing polyphosphate were discovered amongst those samples. GAOs have previously been described as often having tetrad morphology [20]. Two examples of this can be seen amongst the 'Calib. GLY' cells shown here. Occasional tetrads were indeed evident among the samples analysed, however, in accordance with the literature not every glycogen-containing cell had this morphology. Besides the previously known link between tetrad morphology and glycogen content, no other obvious link between 'metabolic state' (the combination of metabolites present) and the morphology of a cell were apparent i.e. it was not possible to predict the contents of a cell based on its appearance. Furthermore, it seems there is no obvious, consistent difference in appearance between cells from different sample sources.

6.4 Automated Identification of Candidate PAOs

Majed *et al.* proposed the identification of functionally relevant populations in EBPR based on the combinations of metabolites present in each individual bacterium [74]. For example, a cell containing polyphosphate with or without other metabolites (PHA, GLY) could be considered to be a candidate PAO. Cells containing glycogen with or without PHAs could be considered to be candidate GAOs. Cells containing just PHAs could belong to either of these populations.

Identifying candidate PAOs based on their Raman spectra could open up new avenues in EBPR research, particularly for the reason Raman analysis is non-destructive and has the potential to be applied to live cells. For example, Raman activated cell sorting [47], at high throughputs [165], could be used to separate live PAOs from other EBPR bacteria for further analysis or in order to highly enrich a culture. This kind of application would require, however, that the Raman analysis is fully automated. The performance of two approaches – univariate (based on peak height data) and multivariate (based on a PLS-DA model) – to automate this analysis are compared here.

6.4.1 Identification of cPAOs Based on Polyphosphate Peak Height

Figure 6-13 shows the results of applying the peak height method to automatically identify polyphosphate in the calibration set of Raman spectra. The spectra and data points are coloured according to whether, by visual inspection, the samples were labelled as containing polyphosphate (blue) or not (green). A polyphosphate-containing sample can be considered to be a candidate PAO (cPAO). The two characteristic polyphosphate peaks are shown in (a) together with the applied peak position boundaries. From the 41 polyphosphate-containing spectra in this data set, the mean peak position of 'polyphosphate peak 1' (PP peak 1) was found to be 697 cm^{-1} with a standard deviation of 1.8 cm^{-1}. The mean peak position of 'polyphosphate peak 2' (PP peak 2) was found to be 1168 cm^{-1} with a standard deviation of 4.9 cm^{-1}. Based on a range defined by four standard deviations (which captures 99.994% of normally distributed data) the peak position boundaries were thus set at 697 ± 8 cm^{-1} and 1168 ± 20 cm^{-1}. For positive identification of polyphosphate in a sample (and therefore indication of a cPAO), Raman peaks with intensities above a specified threshold (set here at three times the noise level) had to be present in both these spectral regions.

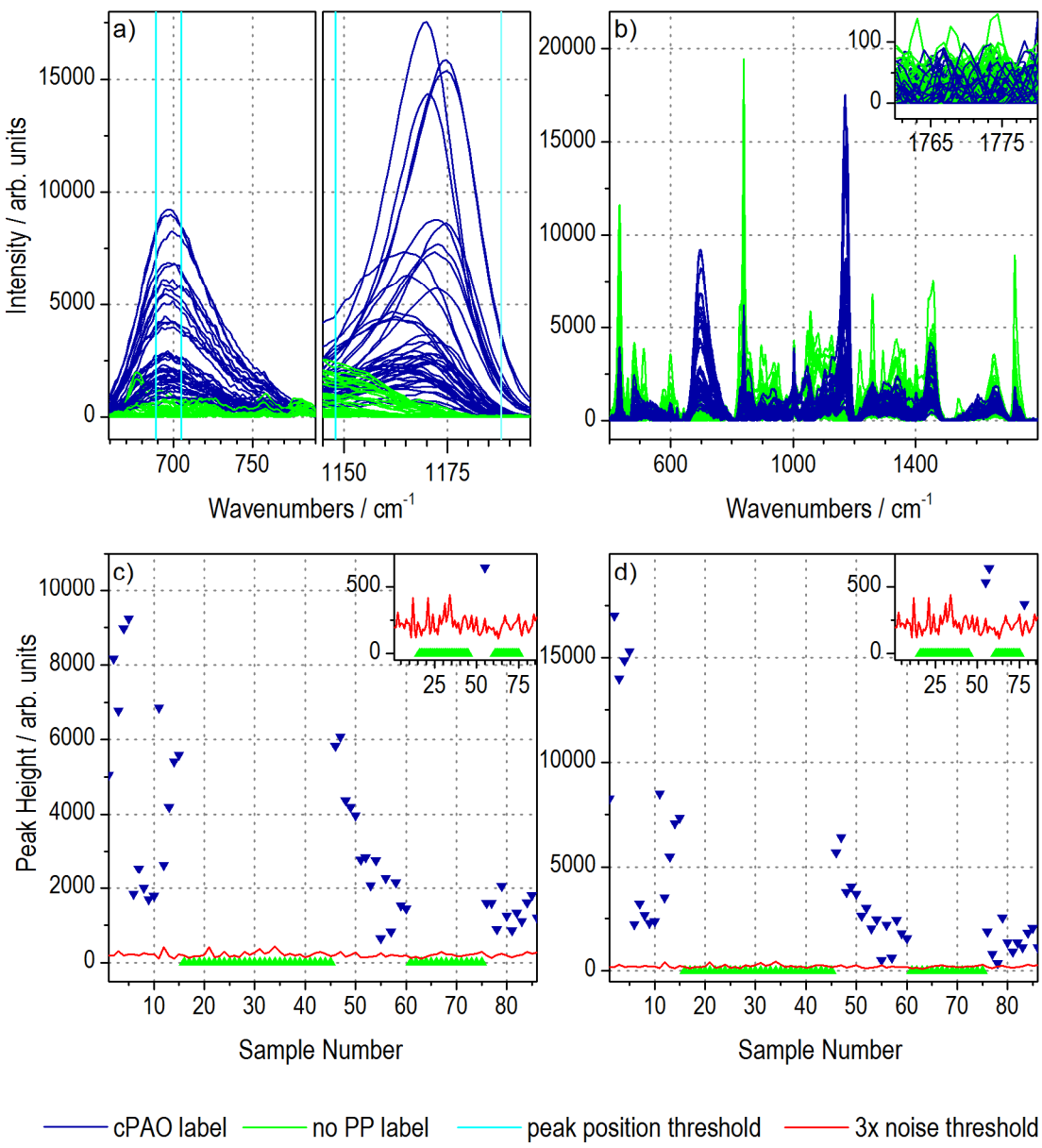


Figure 6-13 Automated identification of PAOs in calibration dataset based on peak height. All the dataset spectra are plotted in (a) and (b). The regions of interest (i.e. polyphosphate peaks) are shown in (a), whereas the full spectra are shown in (b) together with an inset showing the region used to determine the level of spectral noise. Peak heights are plotted in (c) and (d) for PP peak 1 and PP peak 2 respectively. The region near the threshold is shown more clearly in the insets.

Plots (c) and (d) show peak height data for PP peak 1 and PP peak 2 with respect to the intensity threshold. It can be seen that all the samples are classified 'correctly' i.e. in accordance with the presence or absence of polyphosphate as identified by visual inspection

of the spectra. Interestingly, although PP peak 2 characteristically has a greater intensity than PP peak 1, some of the samples for PP peak 2 have intensities closer to the threshold than PP peak 1. This can be explained by the existence of overlapping peaks in spectra where more than one storage compound is present or where the contribution of bacterial components to the spectrum is significant. For example, Figure 6-14 shows Raman spectra for the two samples (55 and 78) with peak heights closest to the threshold. Characteristic Raman bands for PHA and glycogen can be seen in spectrum 55 and 78 respectively. In addition, the spectral contribution from bacterial components is of comparable intensity with those from the storage compounds. The polyphosphate peak at ~ 1168 cm^{-1} can be seen to overlap significantly with other Raman bands in both these spectra, particularly in sample 78 where glycogen is present. Since the peak height determination algorithm used here subtracts a linear baseline drawn between local minima in defined ranges either side of the peak, this affects the relative intensities of the two polyphosphate peaks in the cases where they overlap with other signals. For conservatively determining the presence or absence of polyphosphate in a spectrum, this is not expected to be a problem. For reporting polyphosphate concentration based on the single peak height at ~1168 cm^{-1} as has been done in previous studies [68, 70, 72], however, the difficulty of defining peak height when overlapping bands may be present has been highlighted.

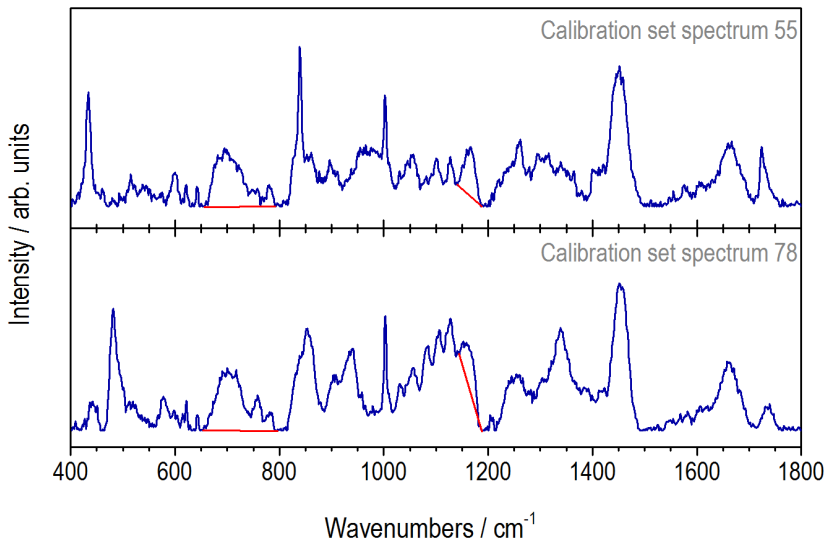


Figure 6-14 Raman spectra numbers 55 and 78 from the calibration data set. Peak-specific linear baselines used by the algorithm to calculate peak height are shown in red.

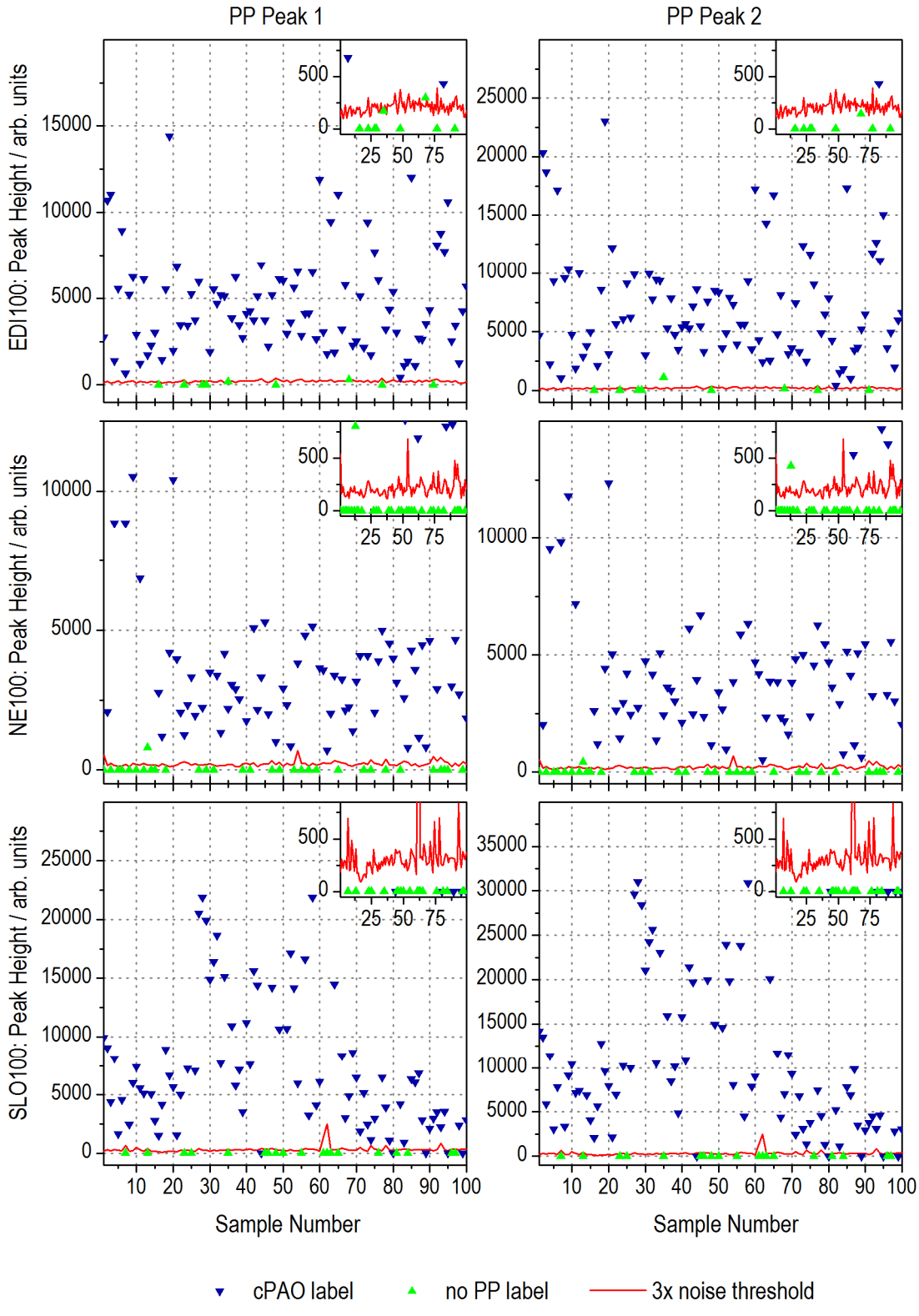


Figure 6-15 Automated identification of PAOs in three test datasets (EDI100, NE100, SLO100) based on peak height. Insets show the region near the threshold more clearly.

Application of the peak height method for identifying PAOs to three further test sets, EDI100, NE100 and SLO100, is shown in Figure 6-15. These data sets represent three different EBPR sources and so provide Raman spectra of bacteria that have come from different microbial communities and been subject to different operating conditions. EDI100 and NE100 come from lab-scale bioreactors whereas SLO100 samples originate from a full-scale EBPR plant. Comparison of spectra from these sources should give an idea of the potential variation in Raman spectra and how this may affect automatic recognition of polyphosphate signals for the identification of cPAOs.

For all three test data sets it can be seen that the majority of samples are classified correctly. Examples of misclassified spectra are present, however, in each data set and some of these are shown in Figure 6-16, together with the micrographs of the bacteria analysed. The appearance of the bacteria in the micrographs are not particularly unusual compared to the other samples studied. Of the 100 samples in EDI100, spectrum 68 is misclassified as containing polyphosphate based on PP peak 1. This spectrum is relatively noisy, with a signal-to-noise (S/N) ratio of 7.4 (based on the phenylalanine signal band at ~1002 cm^{-1} and noise in the region 1760-1780 cm^{-1}) compared to a mean S/N ratio of 27.6 (SD=17.2) for all the spectra analysed. In this spectrum it is hard to judge, therefore, whether the low level signals in the polyphosphate band regions could arise from the presence of polyphosphate. The same is true in NE100 spectrum 13, which was also misclassified by both PP peaks1 and 2. When originally labelling the presence or absence of polyphosphate in these spectra, a conservative decision was made that these samples did not contain polyphosphate. An advantage of automation is highlighted here, in that although polyphosphate has two very characteristic bands, it is not always trivial to decide by inspection if those bands are present or not. Automation introduces a degree of consistency between samples and between analysts in making these kind of judgement calls.

Of the 100 samples in SLO100, five samples (44, 80, 89, 95 and 99) are misclassified as containing no polyphosphate whereas by visual inspection they were labelled as cPAOs. On further inspection it was noted that this classification was based on the position of PP peak 1 lying outside the specified boundaries. Spectrum 44 had an unusual peak shape for which it is hard to tell if polyphosphate is present or not. It appears that the other four spectra could have had another overlapping Raman band present at a slightly higher wavenumber which shifted the apparent PP peak 1 position that was automatically determined. Although some work on the Raman spectra of inorganic polyphosphates has been done, very little is understood about how the band positions and shapes of polyphosphate may change as a

result of interaction with biological components. In order to improve interpretation of these kind of spectra, further investigation of the Raman spectra of polyphosphate with biological components may be beneficial.

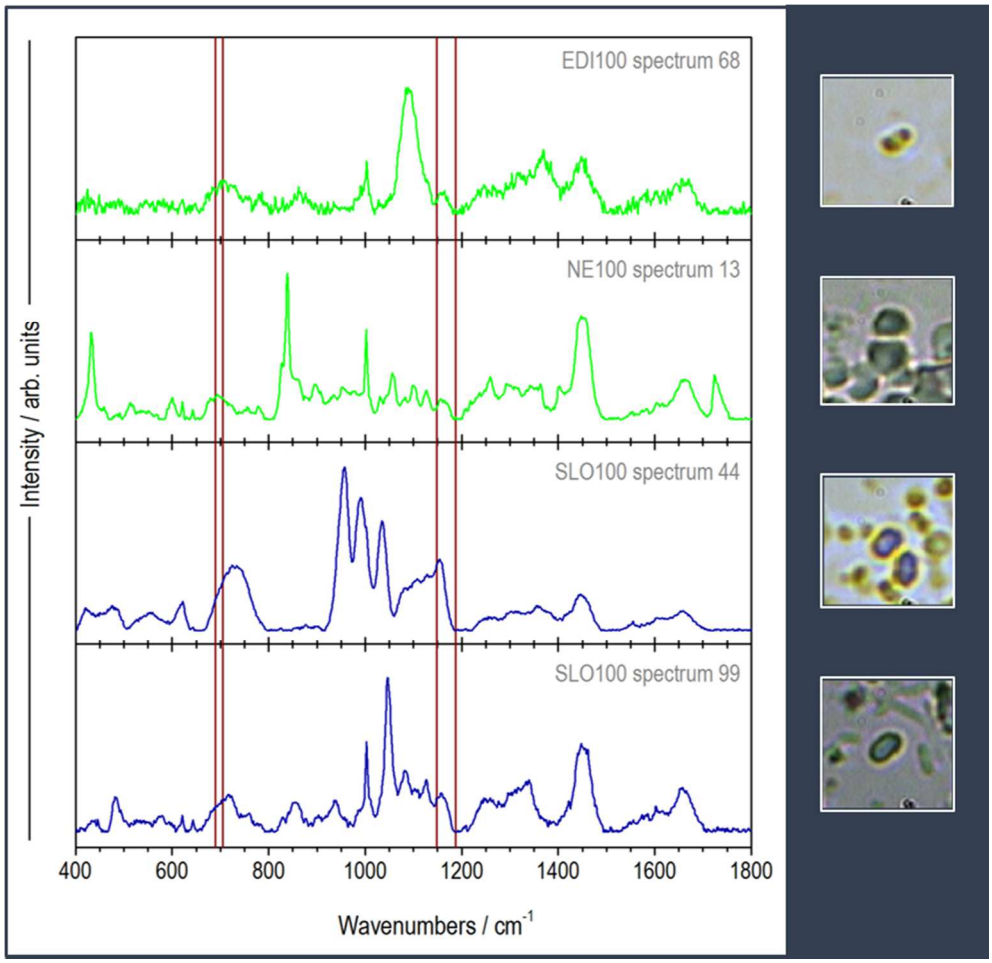


Figure 6-16 Example misclassified spectra from data sets EDI100, NE100 and SLO100. Spectra labelled by visual inspection to contain or not contain polyphosphate are coloured dark blue and green respectively. Peak position boundaries for automated polyphosphate peak recognition and height determination are shown in red.

6.4.2 Identification of cPAOs Based on PLS-DA

Figure 6-17 shows the results of applying PLS-DA to automatically identify polyphosphate (and hence cPAOs) in the calibration set of Raman spectra. Once again, data points are coloured according to whether, by visual inspection, the samples were labelled as containing polyphosphate (blue) or not (green). Based on the RMSECV results and variance captured data shown in (a), the PLS-DA model was built with 3 latent variables. Adding more latent variables can be seen to make little change to the predictive capacity of the model or the variance captured. X-block (spectral) variance is mostly captured in the first two LVs whereas 80% of the Y-block (classification) data variance is represented in the first LV. This is not surprising as the originally assigned classes were based on the presence or absence of polyphosphate and the first LV (subplot (b)) contains predominantly polyphosphate character. Samples scores for the first two LVs are shown in subplot (c). It can be seen that spectra that contain (by visual inspection) one storage polymer cluster well on this plot. Samples containing two polymers, however, are not obviously distinguished from one another by these parameters. The raw classification prediction data from the PLS-DA model is shown in subplot (d). It can be seen that all samples were classified correctly although one sample – number 70 – lies very close to the model-defined threshold. On inspection, it was seen that this spectrum (shown in Figure 6-19) was very similar to SLO100-spectrum 99 (Figure 6-16) and that it may indeed contain polyphosphate unlike its original assignment. The peak height method did not pick up on this apparent inconsistency. The PLS-DA model seems therefore, to be less sensitive than the peak height method to absolute peak position and has the potential to highlight errors made in individual analysis by inspection.

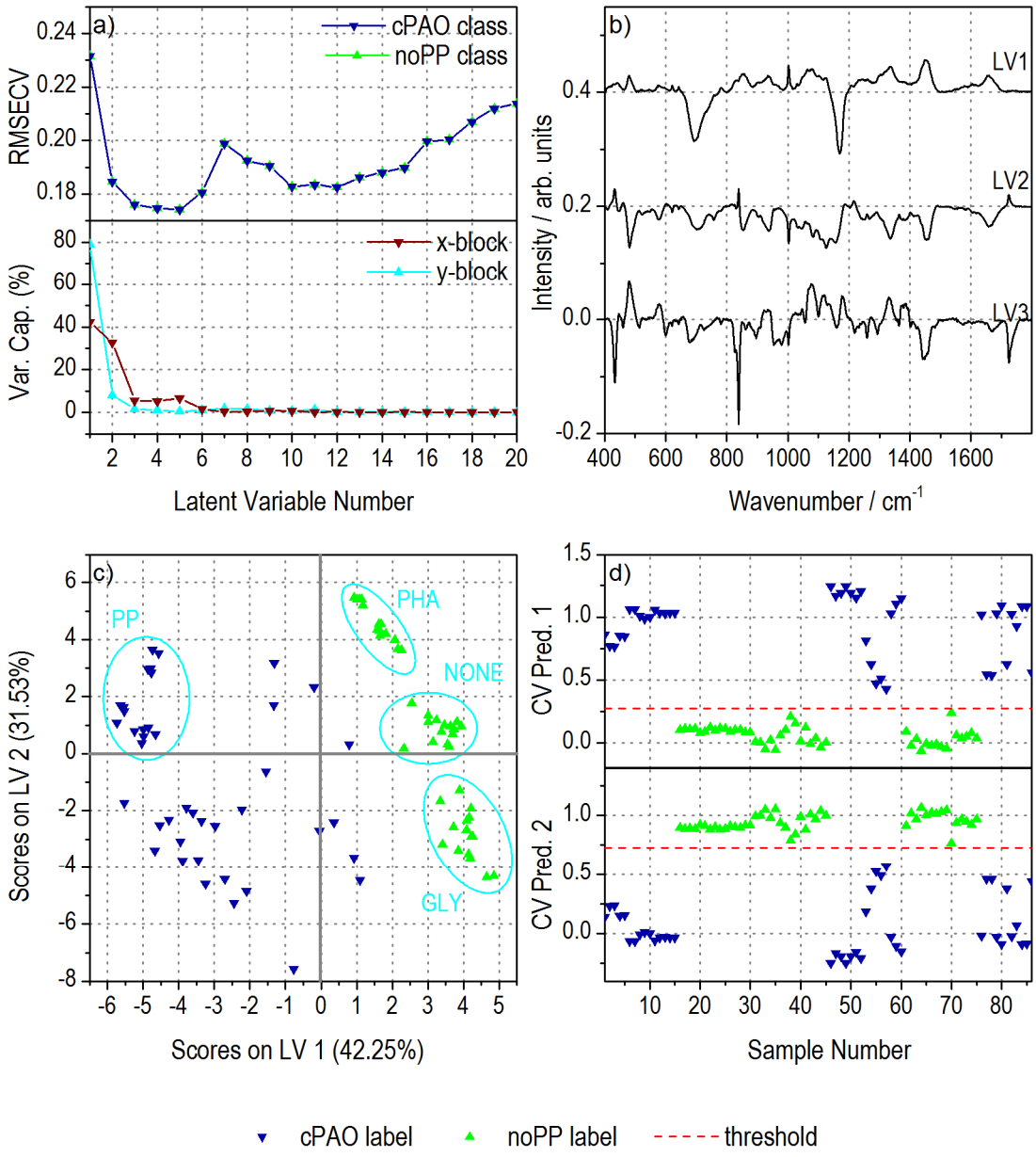


Figure 6-17 Automated identification of PAOs in calibration dataset based on PLS-DA. RMSECV and variance captured per latent variable (LV) number are shown in (a). Loads and scores for the model developed with 3 LVs are shown in (b) and (c). Clustering of spectra containing (by inspection) polyphosphate (PP), polyhydroxyalkanoates (PHA), glycogen (GLY) and no metabolites (NONE) are indicated on the scores plot in (c). The raw cross-validated prediction scores (CV Pred.) for class assignment are shown in (d).

The PLS_DA model built on the calibration data set was applied to the EDI100, NE100 and SLO100 data sets and the results are shown in Figure 6-18. Raw prediction scores for a sample belonging to the 'cPAO' class or the 'no PP' class are shown on the left and right respectively.

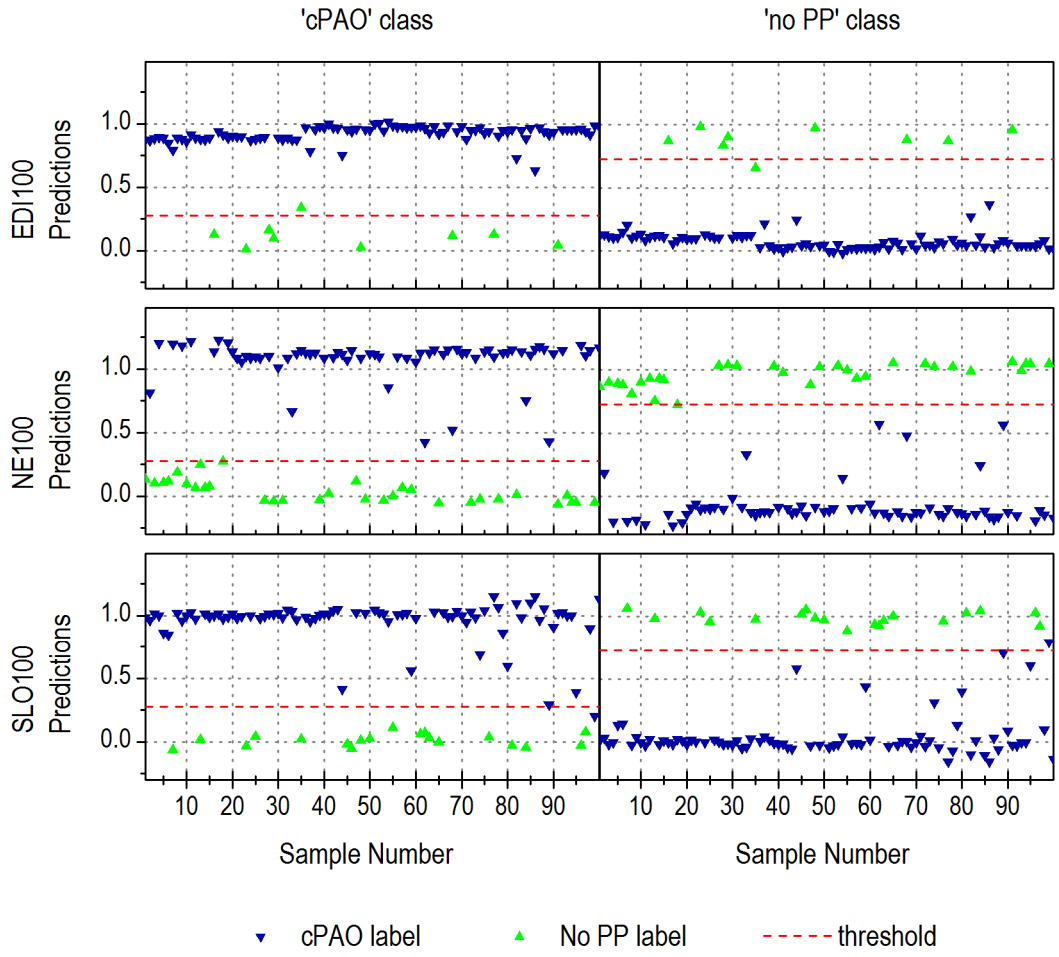


Figure 6-18 Automated identification of PAOs in three test datasets (EDI100, NE100, SLO100) based on the developed PLS-DA model. Raw prediction scores for samples belonging to the 'cPAO' class and 'no PP' class are shown on the left and right respectively.

Amongst these three data sets only two samples were misclassified – one in EDI100 and one in SLO100. In the NE100 dataset, one sample was seen to lie on the threshold for predicting the presence of polyphosphate. The spectra for these three samples are shown in Figure 6-19, together with the micrographs of the bacteria analysed. The appearance of the bacteria in the

micrographs is not particularly unusual compared to the other samples studied. Of the 100 samples in EDI100, spectrum 35 is misclassified as containing polyphosphate rather than not containing polyphosphate. This spectrum is relatively unusual and seems to have a high residual contribution overlapping the position where one of the polyphosphate bands would normally be. This residual may be the reason that this spectrum was misclassified. Of the 100 samples in SLO100, spectrum 99 was the only misclassified sample and this sample has already been discussed in Section 6.4.1 since it was also misclassified by the peak height algorithm. Spectrum 89 from SLO100 is very similar in appearance to spectrum 99 and although it wasn't misclassified, it was positioned close to the threshold.

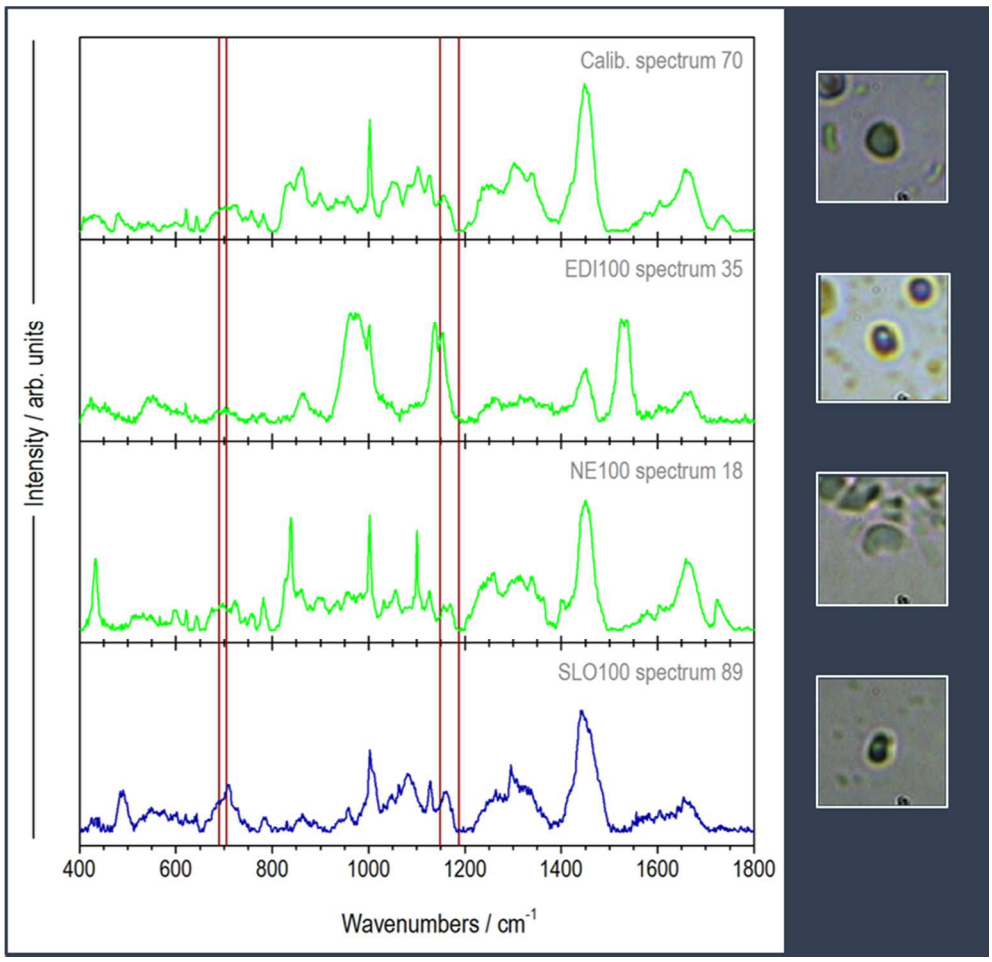


Figure 6-19 Example of a misclassified spectrum from the EDI100 dataset, and examples of spectra for samples lying close to the classification threshold in the calibration, NE100 and SLO100 datasets. Spectra labelled by visual inspection to contain or not contain polyphosphate are coloured dark blue and green respectively. Peak position boundaries from the peak height analysis are shown in red as an aid to identify the major polyphosphate peak regions.

PLS-DA model statistics for all four datasets are shown in Figure 6-20. Hotelling's T-Squared values indicate variance within the model and Q residuals indicate variance outside the model i.e. how much of the spectral data is not fit by the model. The upper 95% confidence interval for normally distributed data is shown on each plot. Statistically, it can be expected that 5% of the data lie above this threshold. If significantly more than 5% of samples lie above this threshold it can be used as an indicator that the model itself may be poor. In this case, 13 samples (3.4%) and 18 samples (4.7%) lie outside the 95% limit for Hotelling's T-Squared and Q residuals respectively. This suggests that this particular PLS-DA model fit the data well. Furthermore, the samples in this upper 5% represented all four data sets and hence suggested that no one data set was fit less well than another by the model.

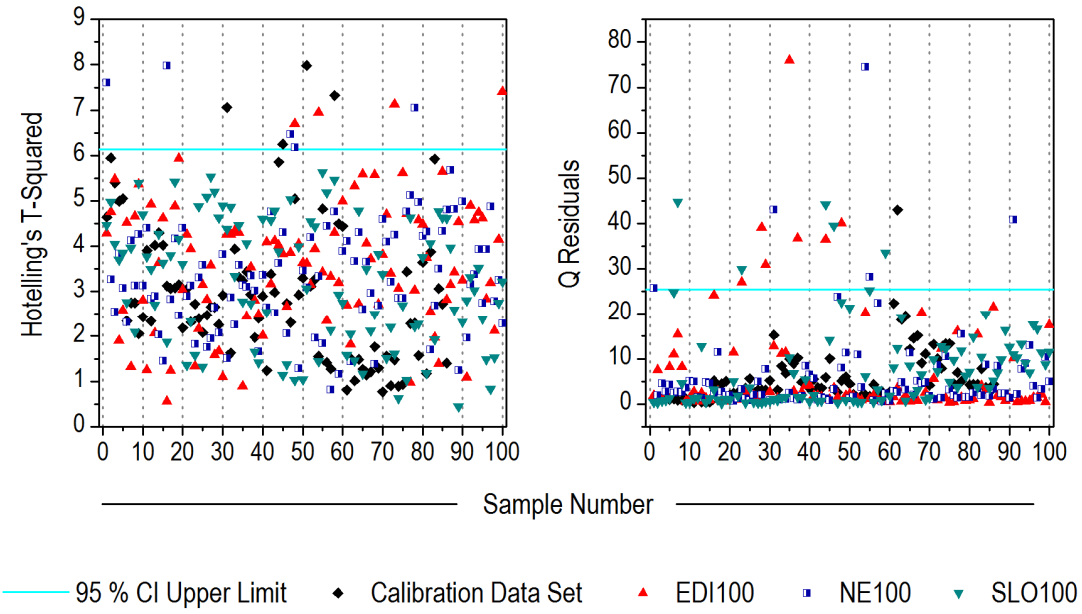


Figure 6-20 PLS-DA model statistics for all four Raman spectra datasets analysed for the presence or absence of polyphosphate.

6.4.3 Comparison of Methods for Automated Identification of cPAOs

A summary of the number of samples misclassified by both automated methods (peak height, PLS-DA) for identification of cPAOs is presented in Table 6-6. The peak height method misclassified slightly more samples (seven) than the PLS-DA method (two) out of a total of 386 samples. The difference in success between the two methods is thus very small when applied to these particular data sets. In addition, there is no gold standard for identifying the presence of polyphosphate in individual bacteria to which the analysis by Raman spectroscopy could be directly compared. So there is no guarantee the original classification of samples by inspection was correct. Overall, therefore, the statistics suggest that both these methods perform similarly well. PLS-DA has the added advantage, however, of the option to analyse residuals in order to easily identify unusual or outlier samples in a large data set.

Table 6-6 Summary of the number of samples misclassified by two automated methods (peak height and PLS-DA) to identify candidate PAOs based on the presence of a polyphosphate signal in the Raman spectra. Both false positive (false +ve) and false negative (false –ve) misclassifications are reported. For the peak height method, results based on each polyphosphate band are reported separately.

Dataset	Peak Height (PP peak 1)		Peak Height (PP peak 2)		PLS-DA	
	False +ve	False -ve	False +ve	False -ve	False +ve	False -ve
Calibration	0	0	0	0	0	0
EDI100	1	0	1	0	1	0
NE100	1	0	1	0	0	0
SLO100	0	5	0	5	0	1
Total	2	5	2	5	1	1

6.5 SUMMARY AND CONCLUSIONS

This study has demonstrated, in agreement with the study by Majed *et al.* [72], that characteristic signals for polyphosphate, PHA and glycogen are clearly identified by Raman spectroscopy when present in individual EBPR bacteria. Building on this, the variability of these signals between samples was investigated in detail for the first time. The position of the major Raman bands for glycogen and PHA were not seen to vary much in spectra of the pure compounds versus spectra of EBPR bacteria. Furthermore, these peak positions were not seen to vary much between the bacterial samples themselves. In contrast, the position of the major polyphosphate Raman bands were shown to vary significantly between pure compound samples and bacterial samples as well as between the bacterial samples themselves. This could raise some challenges for the analysis of EBPR bacteria by Raman spectroscopy. For example, if quantitative analysis of polyphosphate content were desirable, what kind of compound could be used as a standard when a pure polyphosphate sample has such a different signal to that arising from intracellular bacterial polyphosphate? On the other hand, perhaps these peak position shifts could be a useful indicator of some parameter in the cells themselves. This is something that could be considered in future work.

The Raman spectra of EBPR bacteria are complex with many overlapping signals. Despite this, multivariate statistical tools have not yet been applied to these kind of samples. Here, for the first time, two multivariate tools – MCR and PLS-DA – were applied to sample sets of Raman spectra of individual cells obtained from three different EBPR reactors. Using MCR, over 95% of the variance in these Raman spectra could be modelled successfully with four major components. These four components closely resembled pure spectra of the three key EBPR metabolites – polyphosphate, PHA and glycogen – plus a spectrum that could reasonably be considered to represent the underlying bacterial ‘matrix’. For each Raman spectrum (of an individual EBPR bacteria) analysed, a ‘score’ was assigned by the model for each of these components. Plotting these scores gave an overall visual impression of the combination of metabolites present in each sample and the distribution of these among a data set. This may be a useful tool for making quick comparisons of metabolite content in individual cells for large data sets. Furthermore, it would be interesting to compare these plots for samples analysed under different operating conditions. Perhaps they would constitute a kind of ‘fingerprint’ that is characteristic of a particular reactor or operating conditions and may then further be related to process performance. The success shown by MCR in extracting components so similar to the pure compound spectra and the underlying

bacterial matrix is something that may prove useful for quantitative analysis of these spectra. This is investigated in Chapter 7.

The use of PLS-DA for the automatic identification of candidate PAOs based on their Raman spectra was investigated. This kind of analysis could be useful in an application such as high-throughput Raman activated cell sorting – to separate PAOs from the mixed EBPR microbial community so that they can be studied in isolation. For the purpose of the analysis, any bacterium giving rise to a spectrum containing a polyphosphate signal was considered to be a candidate PAO. Compared to a similar analysis made using peak height data, PLS-DA was found to be less sensitive to variation in polyphosphate peak position and misclassified slightly fewer spectra. The option to analyse residuals makes PLS-DA an attractive method with respect to being able to identify samples that are unusual and which should potentially be rejected on that basis. For this kind of application, therefore, PLS-DA is preferred over analysis based on peak height.

Overall, a qualitative analysis of the Raman spectra of a large number of individual EBPR bacteria sampled from different sources has been presented. Furthermore, the application of multivariate tools to these spectra has been demonstrated, such as could open up new avenues of research for EBPR.

Chapter 7
Quantification of Polyphosphate in EBPR bacteria

7.1 INTRODUCTION 165
7.2 METHODOLOGY 171
7.2.1 Normalisation Methods for Quantification of Polyphosphate in EBPR Bacteria 171
7.2.2 Estimation of Limit of Detection and Limit of Identification 174
7.2.3 Spectroscopic Measurements 174
7.2.4 Data Sets 174
7.2.5 Spectral Analysis 175
7.3 PERFORMANCE OF NORMALISATION METHODS 177
7.4 ESTIMATION OF LOD AND LOI OF POLYPHOSPHATE 188
7.4.1 LOD and LOI of Aqueous Polyphosphate Solutions 188
7.4.2 LOD of Polyphosphate in EBPR Bacteria 191
7.5 SUMMARY AND CONCLUSIONS 194

7.1 INTRODUCTION

Polyphosphate is rarely measured directly in EBPR but it could provide a number of valuable insights into PAO behaviour under different operating conditions and hence yield new strategies for improving performance. At the simplest level, measuring the amount of polyphosphate per weight of biomass would yield a useful parameter to use in phosphorus mass balance equations. On a single-cell level, maximum cellular capacities to store polyphosphate could be quantified and enable investigation into how these vary with operating conditions or species. This may help to manage EPBR process conditions or engineer PAOs to allow maximum exploitation of their ability to sequester phosphate from water. Furthermore, new insights may be gained, for example perhaps PAOs also have a minimum, 'recalcitrant' pool of polyphosphate that is not available for use in EBPR. Quantifying all three key EBPR metabolites – polyphosphate, PHA and glycogen – simultaneously could help investigate bacterial stoichiometries and perhaps hypothesised behaviours such as metabolic switching from PAO to GAO type behaviours might be observed directly. The number of studies into methods for releasing and recovering polyphosphate from PAOs is increasing and direct quantification of bacterial polyphosphate could be an invaluable tool for such studies.

Cellular polyphosphate quantification is not only relevant to EBPR, but could be applied in many other fields of research. Polyphosphate is ubiquitous in all biological cells (albeit in much lower concentrations than PAOs) and is thought to be essential for growth and survival [24]. Some pathogenic bacteria such as *Mycobacterium tuberculosis* depend on polyphosphate for their 'virulence and other functions' [166]. In higher organisms, polyphosphate and its biosynthetic enzymes have been shown to be involved in apoptosis in myeloma cells, blood coagulation, osteoblast calcification and breast cancer cell proliferation [167]. The potential for bioremediation of heavy metals such as mercury, copper and uranium using bacteria has been demonstrated, with the ability of these bacteria to sequester heavy metals seeming to be related to the accumulation of polyphosphate [168-172]. Overall, the number of studies on polyphosphate does not seem to reflect its biological and societal importance. One contributing factor to this is thought to be the 'lack of sensitive and facile analytical methods to assess its concentration in biological sources' [24].

A brief summary of existing techniques that may be applied to the quantification of polyphosphate in EBPR bacteria is provided in Table 7-1. Broadly, these include various assays (colourimetric, fluorescence and enzyme), ion chromatography, gel electrophoresis,

and various spectroscopic methods (^{31}P-NMR, ATR-FTIR and Raman). These methods differ from each other on several points of comparison, including (but not limited to) the type of information provided (quantification, cell visualisation, polyphosphate chain-length), the level of the information provided (single-cell or population), the kind of sample pretreatment required (none, cell lysis, chemical extraction), sample size required, time taken for analysis, and the ability of each technique to detect and distinguish polyphosphate in its various bacterial forms.

Bacterial polyphosphates are not uniform macromolecules: They vary in parameters such as composition, arrangement, intracellular location, chain length and solubility. Polyphosphates are polyanions which are stabilised through complexation with metal cations or also, in biological systems, by association with proteins or nucleic acids [173]. As linear polymers, polyphosphates lack inherent tertiary structure, but this chelation with metal ions or association with other biological polymers provides potential for arrangement or templating the shape of the chains in space [174]. Intracellular polyphosphate may be located in the bacterial periplasm or cytoplasm and may be free or associated in different ways with cellular compounds such as proteins and nucleic acids [175]. Seviour *et al.* [11] pointed out that the location of intracellular polyphosphate could reflect different stages in its metabolism, or equally, the differences in cellular organisation of different microorganisms involved in EBPR. Jing *et al.* [176] commented that microorganisms could have multiple stores of polyphosphate that differ in location, chain length and mobility according to the varying functions to which they are applied. The solubility of intracellular polyphosphate depends on both chain length (i.e. the number of phosphate residues) and the nature and concentration of cations to which it is complexed. Solubility may also depend on polydispersity as is demonstrated with the example of Grahams salt – a vitreous sodium polyphosphate with hundreds of residues – that fails to crystallise because of its polydispersity [177]. Variability in solubility leads to the existence of polyphosphate in cells both as soluble complexes of varying chain length, and as insoluble granules. The latter are also known as volutin for the reason that they were first described in *Spirillum volutans* [178].

Table 7-1 Overview of available methods for determination of intracellular polyphosphate.

METHOD *	ADVANTAGES	LIMITATIONS	REFS.
Chemical extraction followed either by direct PP determination (various methods - see other entries in this table) or by PP hydrolysis and determination of OP equivalents 	Widely applied in environmental studies	Difficult to extract all PP leading to underestimation; Low specificity for PP leading to overestimation; PP itself has variable solubility depending on chain length and associated chelating cations; Due to harsh extractants PP may undergo changes during extraction e.g. in chain length, concentration or structure	[75, 173, 179-181]
Basic dye (metachromatic assay) with e.g. toluidine blue 	Relatively simple	Requires extraction; Not specific for PP; Interference from other polyanionic macromolecules e.g. nucleic acids; PP chain length must be >10 for clear metachromic reaction; PP concentration will be underestimated to varying degrees in samples of mixed lengths which contain short fractions	[182-185]
DAPI and quantitative image analysis (QIA) 	No extraction required; Relatively quick; Both short-chain cytosolic PP and PP granules detected; Simultaneous normalisation to biomass possible via comparison with DAPI-nucleic acids complexation	Not yet robust: few studies to date; At higher intracellular PP concentrations (>200 mg/L) linear relationship between image intensity and PP concentration not yet proven	[186, 187]
DAPI fluorescence assay 	No extraction or separation of PP required; High sensitivity; Minimal sample requirement	Salts (conductivity above ~ 1 mS/cm) may suppress fluorescence in short chain (n≤15) PP; Bound (complexed) PP may not be available for binding to DAPI so some sample pretreatment required	[154, 155, 188]
Enzymatic assays 	High specificity; High sensitivity	Requires extraction and separation of PP; Chain length specificity of enzymes may exclude some size classes of PP, particularly short-chain PP	[185, 189, 190]
Ion Chromatography 	Ability to measure chain-length	Requires extraction and separation of PP; High background levels of anions in cell extracts may make it unsuitable for this application; Limited to PP with < 50 residues	[185]

METHOD *	ADVANTAGES	LIMITATIONS	REFS.
Gel Electrophoresis 	Ability to measure chain-length	Requires extraction and separation of PP; Not yet been applied to mixed culture or complex environmental samples	[191-193]
Solution-state ^{31}P-NMR on extracts 	Potential to detect and quantify all P species present; PP separation not necessary; Potential to estimate average chain length of PP; Association of PP with paramagnetic ions (which cause line broadening and signal disappearance) may be removed during extraction process if chelex or EDTA is used	Requires extraction and solubilisation of PP; relatively low sensitivity requiring relatively large sample size; relatively time-consuming	[181, 194]
Solution-state ^{31}P-NMR on suspensions 	Potential to detect and quantify all P species present; PP separation not necessary; May be performed on live cells (i.e. *in vivo*); Potential for dynamic (kinetic) studies; Depending on instrumentation, potential to switch to ^{13}C NMR probe with the same sample to also interrogate carbon species such as acetate, glycogen and PHAs	Only freely mobile pools of PP are visible i.e. insoluble (long-chain) or bound forms of PP are not detected; relatively low sensitivity requiring relatively large sample size; relatively time-consuming; sample heterogeneity causes line broadening; potential for line broadening or signal absence due to association with paramagnetic ions	[176, 186, 195-200]
ATR-FTIR and PLS regression 	No chemical extraction required; Minimal sample requirement; Simple sample preparation; Relatively quick; Simultaneous measurement of multiple analytes	Physical extraction of biotic from abiotic materials may be required if sample matrix interferes due to high absorption by materials such as silica and clay	[201]
Raman 	No chemical extraction required; Minimal sample requirement; Simple or no sample preparation; Simultaneous observation of PP, glycogen and PHA (i.e. physicochemical fingerprinting); Potential to analyse live cells; Potential to visualise and analyse individual cells	Fluorescence background; Peak overlap	[68, 72]

* ATR-FTIR = attenuated total reflectance Fourier transform infrared spectroscopy, DAPI = 4',6-diamidino-2-phenylindole, NMR = nuclear magnetic resonance, OP = orthophosphate, P = phosphorus, PHA = polyhydroxyalkanoates, PLS = partial least squares, PP = polyphosphate, QIA = quantitative image analysis

 visualisation; quantification; chain length; single cell; population; extraction or lysis required

Solution-state ^{31}P-NMR is a powerful technique but can detect only freely mobile pools of polyphosphate (i.e. insoluble, or bound polyphosphate is not detected). Thus, sample extracts are often used in studies where ^{31}P-NMR is employed. Enzymatic assays have high sensitivity and specificity but may not detect short-chain polyphosphates. Likewise, metachromatic assays with dyes such as toluidine blue require chain lengths to be greater than 10 for a clear reaction to be observed. Ion chromatography is one method (along with gel electrophoresis and ^{31}P-NMR) that could actually be used to estimate the chain length of polyphosphates but it is limited to sizes under 50 residues. Furthermore, high concentrations of anions in cell extracts could interfere with this technique to the extent it may not be suitable for this particular application. For seven out of the eleven listed techniques, some kind of sample pretreatment is required, ranging from cell lysis (to release bound polyphosphate) in the case of the DAPI fluorescence assay, to chemical extraction of polyphosphate (for basic dye and enzymatic assays, ion chromatography, gel electrophoresis, and ^{31}P-NMR), to chemical extraction and hydrolysis (colorimetric determination as orthophosphate equivalents). Requirement for chemical extraction is considered to be a notable disadvantage due to the potential for the original form of polyphosphate and other phosphorus-containing compounds to be altered during the process. In addition, recovery of polyphosphate will inevitably be incomplete and extraction efficiency may vary from sample to sample. Raman spectroscopy differs to most other techniques with its ability to quantify polyphosphate on a single-cell level. Population-level data is also potentially available with this method but has not yet been demonstrated for EBPR bacteria. Raman spectroscopy can analyse cells directly and thus no polyphosphate extraction or purification is required. The method is relatively simple and rapid and the fact that it is not targeted at one particular analyte means that many analytes can be investigated simultaneously. Raman spectroscopy, is non-destructive has little interference from water, and may be applied to live cells. Considering all the above, it is clear that Raman spectroscopy has some significant advantages and inhabits a distinct niche amongst existing quantification methods for bacterial polyphosphate.

Raman spectroscopy is theoretically a quantitative technique since the intensity of a Raman spectrum is proportional to the number of scattering molecules being analysed [202]. De Gelder *et al.*, for example, used a Raman band at ~1734 cm^{-1} to monitor PHB production in *Cupriavidus necator* DSM 428 (H16) [65]. Practically, however, using Raman spectroscopy for quantitative analysis can be complex since the intensity of a spectrum is also affected by instrumental and sample-related factors including laser wavelength and power (which are

often unstable), optical alignment, sample presentation, fluorescence, and variation in analyte peak shape or position [203]. Operating the Raman instrument in a temperature-controlled room would improve the stability of laser wavelength and power by removing temperature-dependent factors [204], however, the laser diode itself also has inherent instability [205]. In the only previous application of Raman spectroscopy to quantify polyphosphate in EBPR bacteria (furthermore, in any bacteria) [68, 72], the intensity (as peak height) of one of the characteristic bands of polyphosphate (at 1168-1177cm^{-1}) was reported directly in CCD counts. The representation of intensity as peak height rather than peak area in this case is justifiable due to the presence of many overlapping bands in these Raman spectra. Reporting absolute intensity can occasionally be sufficient for data comparison in the case that a single data set is measured on a single instrument in a short period of time relative to the stability of the instrument [202]. To render Raman-based quantification of polyphosphate more universally comparable, however, some kind of spectral normalisation is vital. In this study we compare for the first time several options for normalising the Raman spectra of EBPR bacteria for the purpose of quantifying polyphosphate. Four methods are compared: two common methods based on peak height ratios and two novel methods based on MCR analysis.

A second aim of this study is to estimate the limit of detection (LOD) for Raman-based quantification of polyphosphate. Limit of detection for a particular technique refers to the threshold under which the analyte of interest cannot be reliably detected. This parameter is of vital importance for data interpretation. For example, *Pseudomonas putida* and *Acinetobacter johnsonii* are two types of bacteria known to accumulate large amounts of polyphosphate under certain conditions. Raman spectra of dried cell pellet smears of these bacteria made by this author, however, yielded no characteristic polyphosphate bands. Without having an appreciation of the technique LOD it is hard to assess whether the reason for this was that there was no polyphosphate present in the bacteria, or if the growth conditions or experimental parameters were such that any polyphosphate present could not be detected. Likewise, in the Raman-based identification of candidate PAOs in Chapter 6, it would be useful to know the threshold under which polyphosphate cannot be measured and thus how many bacteria which could be PAOs are not being detected as such. Thus, an estimation of the LOD for cellular polyphosphate quantification with Raman spectroscopy would prove useful in data interpretation and also in comparing and selecting methods for any particular application.

7.2 METHODOLOGY

7.2.1 Normalisation Methods for Quantification of Polyphosphate in EBPR Bacteria

Using peak height ratios rather than absolute peak heights is a simple, commonly used approach to normalising Raman spectra. This requires the existence of a reference band within all the sample spectra that does not interfere with the characteristic analyte bands and which does not change from sample to sample. In biological spectra, proposed reference bands have included the breathing mode of phenylalanine at ~1002 cm^{-1}, the C-H_2 deformation band at 1440-1460 cm^{-1} and the amide I band at 1650-1680 cm^{-1} [206]. These options can be considered for the particular case of intracellular polyphosphate quantification in EBPR bacteria.

Figure 7-1 shows (in solid blue) a typical Raman spectrum for a bacterium. This is overlaid (in orange, green and cyan respectively) with Raman spectra acquired from pure compounds of SPG45, glycogen and polyhydroxybutyrate, which represent the three key metabolites in EBPR. As discussed in Chapter 5, when these metabolites are present in the analysed bacteria, they can give rise to strong bands that in some cases dominate the Raman spectra. The position of the candidate reference bands as discussed above, are shown in red to highlight where they may overlap with other variable bands in the spectra. For example, it can be seen that the C-H_2 deformation band at 1440-1460 cm^{-1} overlaps with significant peaks arising from both glycogen and PHB. Since the sample bacteria contain variable amounts of these components, this particular band does not meet the criterion that a reference band should not vary between spectra. The phenylalanine peak at ~1002 cm^{-1} is not an ideal reference band either since it has a small overlap with the edge of a glycogen peak. It can be seen that the amide I band at 1650-1680 cm^{-1} does not overlap with bands from any of the other three spectra. From the three candidate reference peaks, it can therefore be considered as the best option.

Applying the same examination to the two major polyphosphate peaks at ~698 cm^{-1} and ~1172 cm^{-1}, it can be seen that both peaks overlap to some extent with bands from the other three components (PHB, glycogen and bacterial contributions). In particular the polyphosphate band at ~1172 cm^{-1}, although previously being used as a proxy for polyphosphate concentration in EBPR bacteria, overlaps with the edge of a major glycogen band. Thus it may be anticipated that polyphosphate measurements based on peak height

measurements are not strictly comparable when the intracellular quantities of PHB and glycogen also vary between samples.

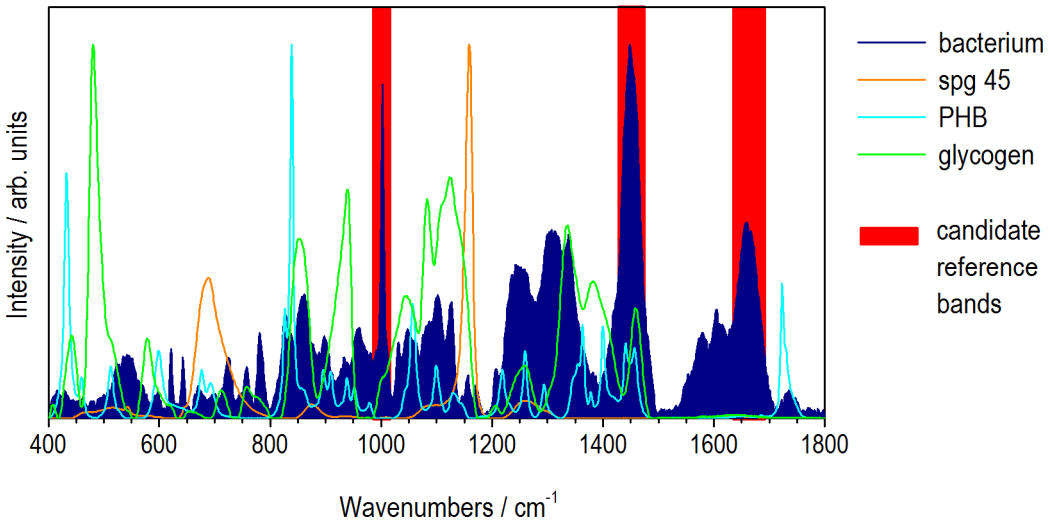


Figure 7-1 Spectra of three pure compounds (sodium phosphate glass type 45 – 'spg45'), PHB and glycogen) representing key EBPR metabolites and a typical bacterium spectrum, each normalised to the peak height of the most intense spectral band. Previously suggested bands to use as internal standards for normalisation of biological Raman spectra are highlighted in red.

An alternative method to using peak height data for normalisation and quantification is to use information contained in the entire Raman spectra i.e. to take a multivariate approach. In Chapter 6, the application of a type of multivariate analysis – MCR – to the Raman spectra of EBPR bacteria was demonstrated. It was shown that a 4-component MCR model yielded loadings which were very similar to the three pure metabolite spectra (for polyphosphate, PHA and glycogen) and the underlying bacterial spectrum. We consider here that this kind of model may be useful for normalising polyphosphate measurement.

Figure 7-2 shows two examples of EBPR bacteria spectra overlaid with components 1 (representing polyphosphate character) and 4 (representing bacterial character) from a 4-component MCR model applied to each spectrum. It can be seen from this that scores on the polyphosphate-type loading could potentially be used as a proxy for polyphosphate concentration, whereas scores on the bacterial-type loading could be used as a reference for normalisation.

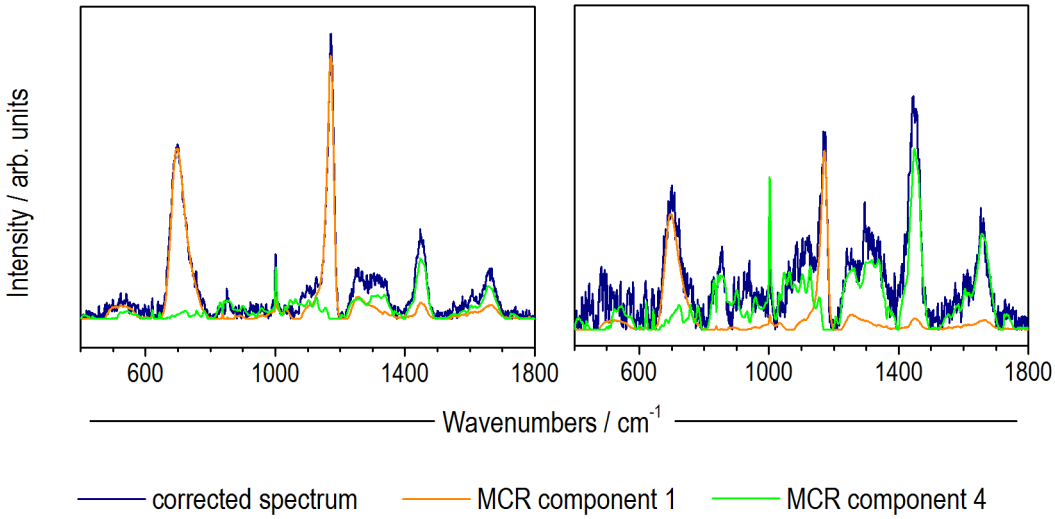


Figure 7-2 Two examples of EBPR bacterial spectra overlaid with components 1 (representing polyphosphate character) and 4 (representing bacterial character) predicted using a 4-component MCR model built on the calibration data set used in Chapter 6.

The principal advantages of using multivariate calibration rather than univariate calibration for regression models in analytical chemistry has been discussed by Bro [207]. Many of the principles discussed therein can be applied to the normalisation problem faced here. In short, the multivariate approach is less sensitive to spectral noise and has the power to overcome the problems of overlapping bands previously discussed. Furthermore, the peak height method has no power to indicate any potential errors in the reported results arising from unusual spectra or interferents since it relies on data from just one or two data points. By contrast, the residuals (spectral components that do not fit) in a multivariate model constitute a powerful indicator of samples for which a measurement may not be reliable.

In this study we compare the performance of four normalisation methods (two peak-height-ratio and two multivariate) on a test set of 400 spectra. The peak-height-ratio methods use phenylalanine and amide I bands respectively for normalisation. The multivariate methods apply two different MCR models applied to the data, from which the scores on the bacterial-type loading are used to normalise the scores on the polyphosphate-type loading. These models are described in Section 7.2.5.

7.2.2 Estimation of Limit of Detection and Limit of Identification

Limit of detection (LOD) for a particular technique refers to the threshold under which the analyte of interest cannot be reliably detected. For pure component spectra this could relate to the intensity at which the most prominent band in the Raman spectra becomes distinguishable from the spectral noise. In more complex mixtures containing unknown contributors, however, a limit of identification (LOI) may also be defined at which several bands from the analyte of interest are distinguishable from noise [208]. This means that presence of the analyte is unmistakeably identified and cannot be confused with a signal arising from another source. For polyphosphate only two strong characteristic bands are present in the Raman spectra. Thus the LOI is defined here as the intensity at which the lower intensity band (at ~693 cm^{-1}) becomes distinguishable from spectral noise.

Estimation of LOD (and LOI) requires that polyphosphate 'standards' of known concentration are measured. Ideally these would be in the same form as the samples to be analysed. However, bacterial cells containing known quantities of polyphosphate do not exist and so at present, a pure compound must be used instead. Here we used two commercially available 1 M (with respect to phosphate) polyphosphate solutions from Kerafast, Inc (MA, US). Raman spectra were acquired of various dilutions of these samples to give an initial estimation of LOD and LOI. Since the signal-to-noise ratio of Raman spectra increases with acquisition time, LOD and LOI were determined for the shortest and longest acquisition times used in this study: 10 and 200 second extended scans respectively.

7.2.3 Spectroscopic Measurements

Raman spectra were acquired with the instrumentation and the set-up described in Chapter 4. Bacterial samples and aqueous polyphosphate samples were also prepared according to the methods described in Chapter 4. Point spectra of bacterial and aqueous polyphosphate samples were acquired over the range 400 to 1800 cm^{-1}.

7.2.4 Data Sets

For investigation of normalisation methods, a data set of 400 Raman spectra was built from mixed liquor samples taken from SBR-D during an uptake-release test. The test was performed in a similar manner to the batch tests described in [209]. Briefly, at the end of a process cycle, the reactor contents have been halved to 1 L since 1 L of effluent has been drawn off. At this point, 100 mg/L COD as sodium acetate was added to the reactor. Subsequently, 45 minutes anaerobic phase with nitrogen purging was followed by 195

minutes aerobic phase. Aqueous orthophosphate, COD, temperature, pH and dissolved oxygen concentrations were monitored throughout the test. TSS and VSS in the reactor during the test were determined to be 5.94 g/L and 3.74 g/L respectively. One sample for Raman spectroscopy was taken immediately prior to beginning the test, and seven samples were taken through the test – after 7, 18, 45, 60, 90, 150 and 240 minutes. The first four samples were taken during the anaerobic phase of the test, and the last four samples were taken during the aerobic phase. For each of the eight samples, Raman spectra were acquired from 50 cells, each with a 10 second extended scan. All 400 spectra were acquired within a day.

The calibration data set used in multivariate analysis of the results was the same base set as that used in Chapter 6. Briefly, this comprised 86 Raman spectra of EBPR bacteria sampled from two 4-litre SBRs located in the research group of Professor Gu at the Department of Civil and Environmental Engineering, Northeastern University, M.A. The spectra were selected on the basis that they covered all the metabolic states that were encountered in spectra acquired in this work. In addition, the calibration data set included 8 spectra each of PHB and glycogen for the purpose of aiding the MCR algorithm to find a solution for the pure component spectra in the samples.

For estimating polyphosphate LOD, two data sets of 23 Raman spectra were built from analysing aqueous solutions of polyphosphate with varying concentration. The two data sets differed in acquisition time of the Raman spectra – selected to be 10 seconds for one set and 200 seconds for the other. ‘High molecular weight’ and ‘medium chain length’ polyphosphate solutions were purchased from Kerafast (Boston, MA). The reported average chain lengths for these two solutions were 700 and 100 phosphate units respectively. The reported concentration for both solutions was 1 M (with respect to phosphate units). Using ultrapure water, a dilution series of each sample was prepared with concentrations ranging from 0.05 M to 1 M. This was repeated on a later day so that in total four sets of Raman spectra were collected.

7.2.5 Spectral Analysis

Spectral Preprocessing

Raman spectra of EBPR bacteria and aqueous polyphosphate samples were preprocessed using MATLAB to remove background and baseline – using the orthogonal vector

correction method described by Maquelin *et al.* [141] and the SWiMA method proposed by Schulze *et al.* [123] respectively.

Normalisation Methods

Peak heights for polyphosphate peaks 1 and 2 (PP1 and PP2), phenylalanine (PHE) and amide I (AMI) bands (at approximately 697, 1172, 1002, and 1657 cm^{-1} respectively) were determined in MATLAB using the 'max' function in specified regions of the spectra.

Two 4-component MCR models for determination of polyphosphate from Raman spectra were built using PLS_Toolbox. These are described below:

MCR model 1 (MCR1) was built on the calibration dataset described in Section 7.2.4. No preprocessing (after background and baseline correction) was applied to the calibration data since it was desired that the spectra of the pure compounds (which were more intense than the other samples) would have high leverage in building the model. It was hoped that this would cause the model loadings to be as similar as possible to the real underlying metabolites. Model statistics indicated that this model fit 99.3% of the variance observed in the calibration spectra.

MCR model 2 (MCR2) was very similar to MCR model 1 except that the calibration data set included 8 repeats of a synthetic polyphosphate-type spectrum to match the 8 spectra each of pure glycogen and pure PHB which were already present in the dataset. This synthetic polyphosphate spectrum was constructed from the polyphosphate-type component yielded in MCR model 1. The two major polyphosphate peaks were selected from this component spectrum and every other data point in the new spectrum was assigned to zero. This synthetic spectrum was multiplied by 250000 to make its intensity comparable with the two other pure metabolite spectra and so that it would also have high leverage amongst the samples for building the model. No preprocessing was applied to the calibration data (after background and baseline correction). Model statistics indicated that this model fit 99.4% of the variance observed in the calibration spectra.

Limit of Detection of Polyphosphate

The 1-component PLS model for estimation of LOD of polyphosphate was built using PLS_Toolbox. No preprocessing was applied to the spectra besides the background and baseline removal technique already described. For the Raman spectra acquired with 10 second extended scans, the PLS model was found to capture 90.7 of X-block variance and 97.0 % of Y-block variance. For the Raman spectra acquired with 200 second extended

scans, the PLS model was found to capture 97.2 % of X block variance and 98.8 % of Y-block variance.

Atomic Force Microscopy (AFM) was used to estimate how much of the Raman laser spot might be filled by our bacterial samples. AFM was carried out under ambient conditions with a commercial scanning probe microscope: the Explorer AFM (Veeco). AFM images were acquired using silicon nitride cantilevers (Bruker MLCT-B) with a nominal tip radius of 20 nm. Non-contact mode imaging was employed to minimise sample disturbance. A 50 μm square region was imaged at a scan rate of 1 Hz giving a tip scan speed of 50 μm/s. The pixel density was 400 per scan line producing a lateral scan resolution of 125 nm. Vertical resolution is influenced by a range of imaging parameters; in this instance it can be asserted to be < 1 nm and hence substantially less than the vertical scan range (~ 500 nm). These parameters can thus be considered negligible with respect to the imaged bacteria.

7.3 Performance of Normalisation Methods

Samples for comparing normalisation methods were taken from an uptake-release test as described in Section 7.2.4. Monitoring data for this test are shown in Figure 7-3. It can be seen that orthophosphate and COD concentrations in the mixed liquor follow the expected profile throughout the test: During the first 45 minutes (the anaerobic phase), the orthophosphate concentration (measured as PO_4-P) rises to a maximum of 85 mg/L, and the 100 mg/L COD added (in the form of acetate) at the start of the test is concurrently consumed during this period. The measured COD had a minimum concentration of 20 mg/L despite the expectation that any residual COD after the anaerobic phase would be rapidly consumed during the aerobic phase. It was concluded this was not due to particulate COD since a comparison of centrifuged and syringe-filtered (0.45 μm pore size) samples yielded the same results. Whether this was due to residual COD such as that described by Gaudy and Blachly [210] or whether it was the result of a systematic experimental measurement error requires further investigation. In any case, the COD profile partially follows the expected trend by rapidly decreasing to a minimum within the anaerobic phase. Temperature in the reactor varied from 21.4 °C at the start of the test to 22.9°C at the end of the test. DO concentrations during the aerobic phase (between 45 and 240 minutes) fluctuated due to manual switching on and off of the aeration in an attempt to maintain a target concentration of 4 mg/L. pH varied between 7.0 and 8.1 over the test, and a small aliquot of 5 M HCl was added 120 minutes into the test to keep the pH in the target range of 7.0 to 8.0.

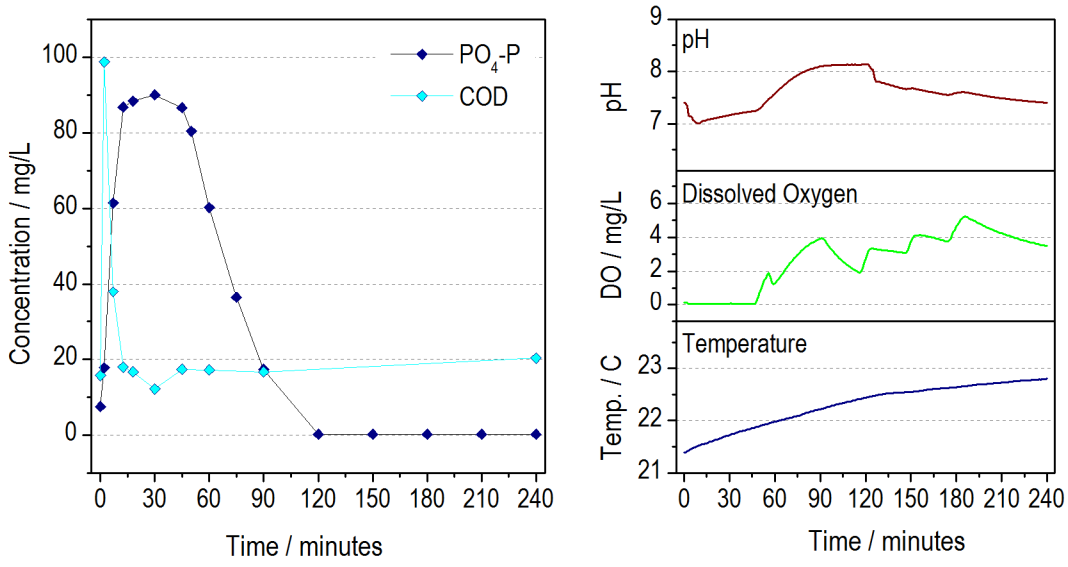


Figure 7-3 SBR monitoring data acquired during the uptake-release test from which samples were taken for Raman spectroscopy.

Eight samples were taken from the reactor for Raman spectroscopy during the uptake-release test. Fifty cells from each of these samples were analysed with Raman spectroscopy. All 400 of these spectra were acquired as one batch within one day. This helps to minimise variation within the data set of laser wavelength and power, temperature and optical alignment (since no instrument adjustments were made between samples). Quantitative analysis using absolute Raman intensity is often successful for such single data sets acquired in these kind of conditions [202]. The 'direct' (i.e. not normalised) intensities of peak heights in this data set may thus be considered to provide a reasonable point of comparison for the normalisation methods which were trialled.

Four normalisation methods were applied to the dataset. These comprised two peak height methods: the ratio of the height of PP peak 2 (at ~ 1172 cm^{-1}) to the phenylalanine peak at ~1002 cm^{-1} and the ratio of the height of PP peak 2 to the amide I band at 1657 cm^{-1}. Two MCR-based approaches to normalisation were also used (termed MCR1 and MCR2). The MCR models were built as described in Section 7.2.5 and for each of these models the scores on the bacterial-type component were used to normalise the scores on the PP-type component. The loads for both of these models are shown in Figure 7-4.

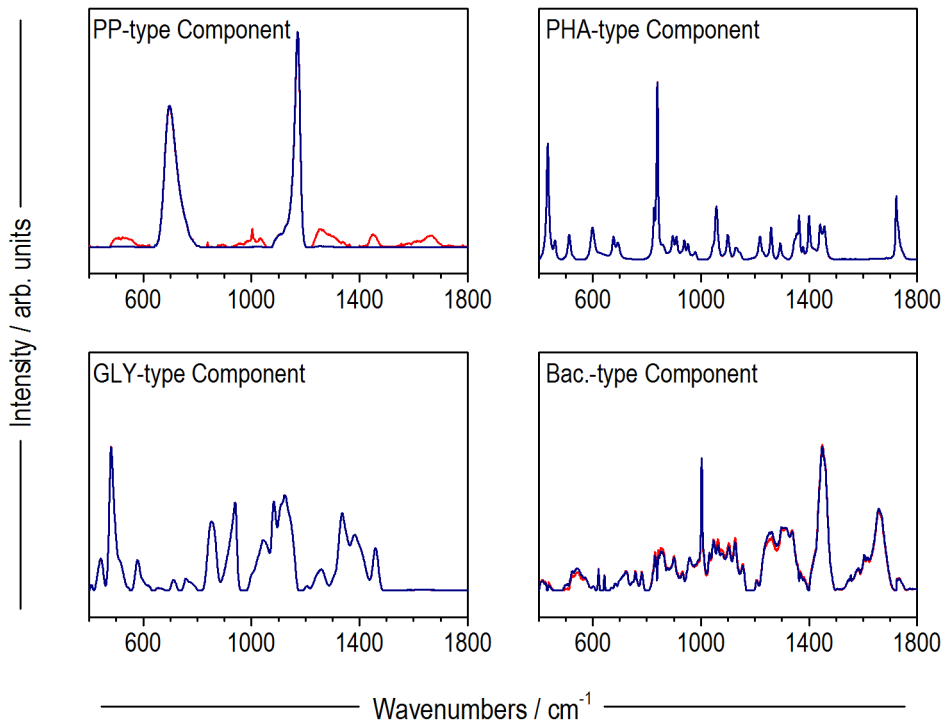


Figure 7-4 Loadings for MCR model 1 (shown in red) and MCR model 2 (shown in blue). The loadings for the PHA-component and GLY-type component are the same so only one of the two plots is visible.

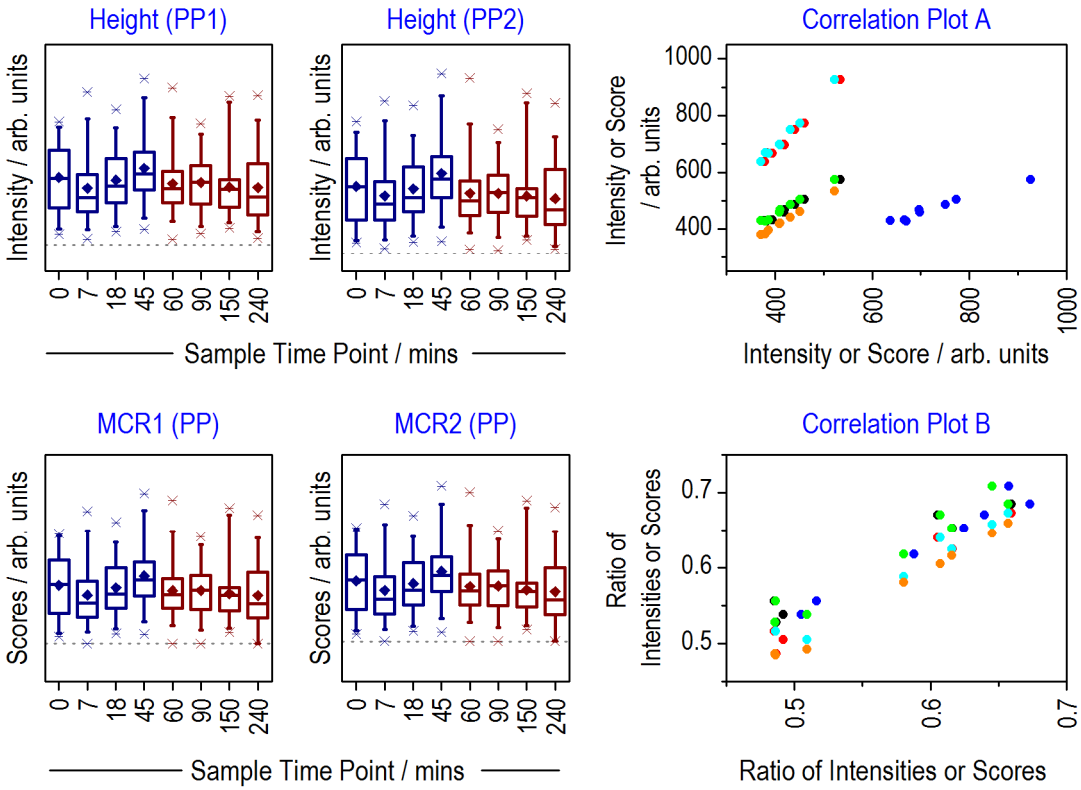
It can be seen that the loadings for the 4 components are very similar for both MCR models. The most significant difference, as intended, is that the PP-type component in model 2 has deliberately been engineered (biased) to remove any spectral contributions that could arise from bacterial contributions. Hence, in contrast to the PP-type component from MCR1, the component in MCR2 has no peaks evident in regions such as ~1002, 1450, and 1660 cm^{-1}. Loadings for the PHA- and GLY-type components are exactly the same, likely because they are influenced so heavily by the inclusion of pure PHA and GLY spectra in the calibration dataset. Loadings for the bacteria-type (Bac.-type) component for MCR1 and MCR2 are very similar.

The results of the four normalisation methods applied to the Raman spectra are shown in Figure 7-5, Figure 7-6 and Figure 7-7. The 'direct' (i.e. not normalised) measurements for polyphosphate are shown in Figure 7-5. These correspond to peak height for the two characteristic polyphosphate peaks (PP1 at ~ 697 cm^{-1} and PP2 at ~ 1172 cm^{-1}), and scores on the PP-type component for both MCR models. The 'reference' measurements used for normalisation are shown in Figure 7-6. These comprise the peak heights of characteristic

phenylalanine and amide I bands (at ~1002 and ~1657 cm^{-1} respectively), and scores on the bacterial-type component for both MCR models. The ratios of the 'direct' measurements to the reference measurements i.e. the normalised measurements are shown in Figure 7-7.

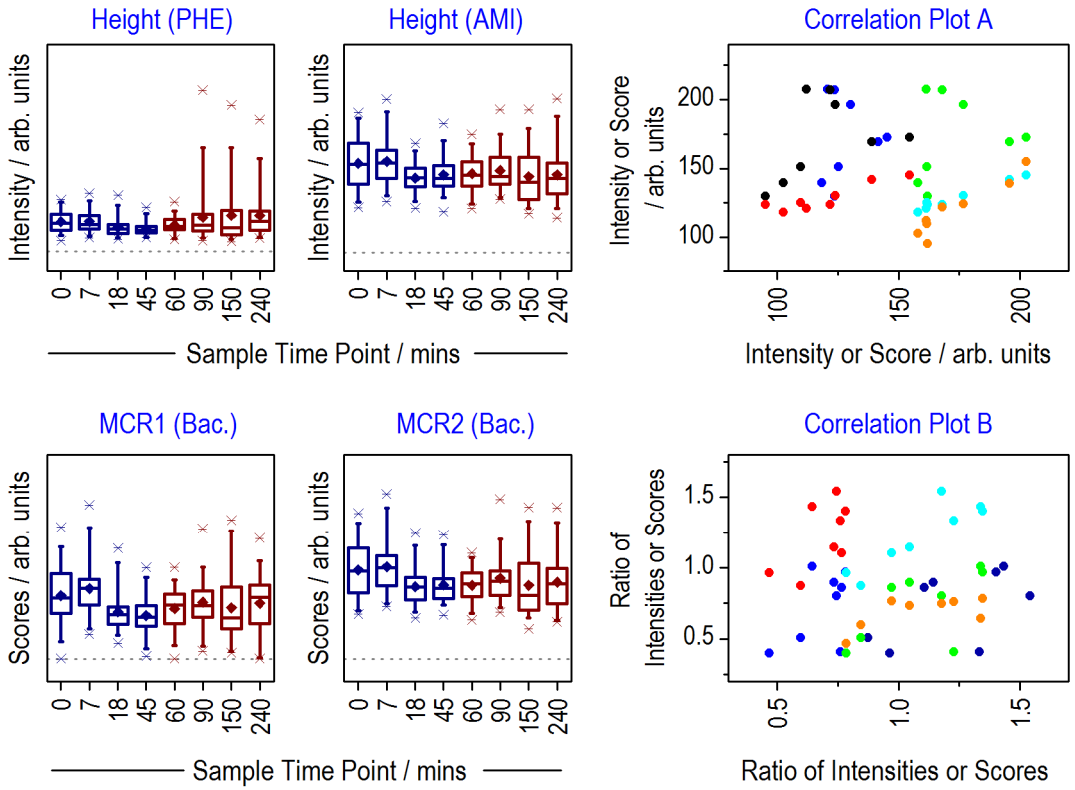
The format of these figures is that they present box plots, correlation plots and a table of correlation coefficients. On each box plot, eight samples are shown that correspond to the eight mixed liquor samples taken from the SBR during the uptake-release test. The first four of these (shown in blue) were taken during the anaerobic phase and the last four (shown in red) were taken during the aerobic phase. The statistics indicated by each sample box represent 50 Raman spectral measurements made on individual EBPR bacteria. The box itself encapsulates data from the 25^{th} to 75^{th} percentile and the whiskers show the limits of the 5^{th} to 95^{th} percentile. The stars indicate outliers. The line in the centre of the box represents the median value and the diamond symbol for each sample represents the mean. The dotted grey line indicates a value of zero intensity or zero score. Correlation plot A shows paired data (i.e. comparison between pairs of measurements) for the mean of each set of 50 spectra taken per sample. It should be noted that on these plots the MCR model scores were scaled down by a factor of 10 so they were in the same order as peak heights measured directly. Correlation plot B shows paired data for the ratio of the mean to the range (5^{th} to 95^{th} percentile) of each set of 50 spectra. This compares methods with respect to the distribution of data within each set of 50 spectra. The table in each figure presents the key for the six pairs of data compared on the correlation plots together with the associated correlation coefficients.

Figure 7-5 shows that the direct (not normalised) measurements of both polyphosphate peak heights and both MCR models are very similar. This is indicated by the fact that the correlation coefficients of the data in both plot A (mean for each sample) and plot B (mean divided by range for each sample) are greater than 0.9 for all of the paired data sets.



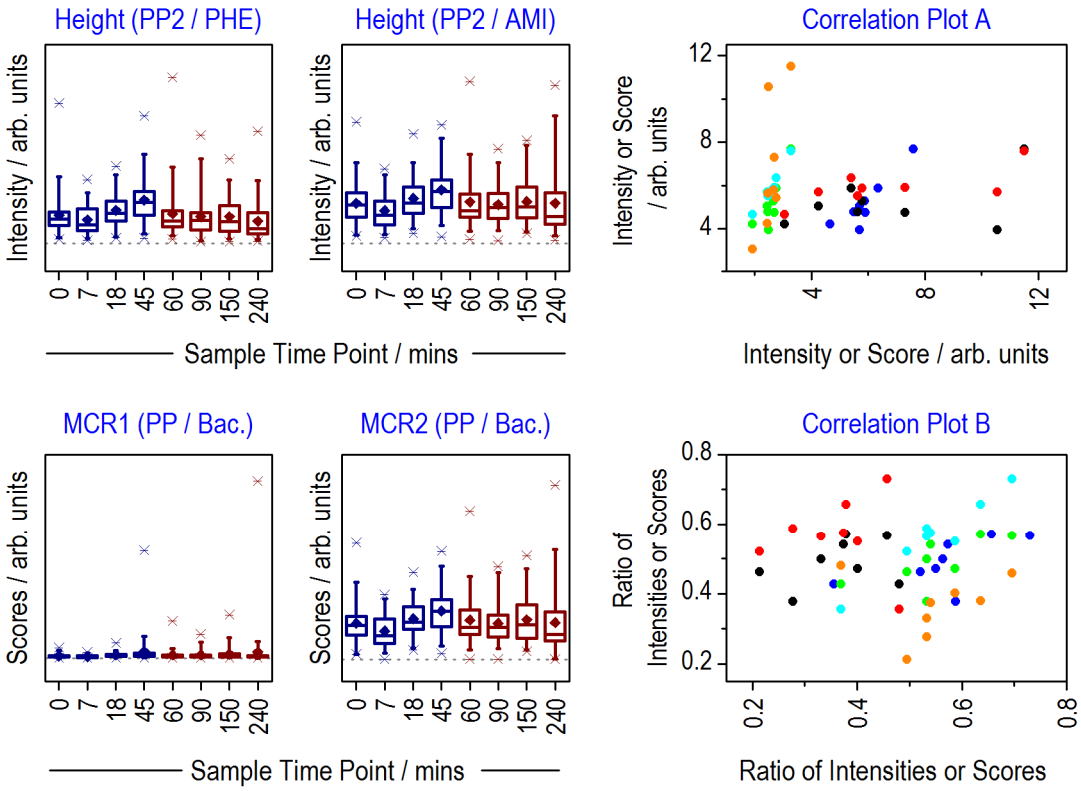
Symbol	Correlation Pair	Correlation Coefficient (linear fit)	
		Plot A	Plot B
blue •	Height (PP1) / Height (PP2)	0.968	0.977
black •	Height (PP1) / MCR1 (PP)	0.993	0.951
green •	Height (PP1) / MCR2 (PP)	0.991	0.945
red •	Height (PP2) / MCR1 (PP)	0.981	0.974
cyan •	Height (PP2) / MCR2 (PP)	0.987	0.970
orange •	MCR1 (PP) / MCR2 (PP)	0.999	0.994

Figure 7-5 Four candidate 'direct' (i.e. not normalised) measurements for determination of polyphosphate using Raman spectra. These comprise peak height for the two characteristic polyphosphate peaks (PP1 at ~697 cm-1 and PP2 at ~1172 cm-1, and scores on the PP-type component for both MCR models. Paired correlation plots are shown for direct comparison between the methods. Correlation plot A compares the mean of each sample (representing 50 spectra) and correlation plot B compares the ratio of the mean to the range (5^{th} to 95^{th} percentile) for each sample. Further explanation of the plots are given in the text.



Symbol	Correlation Pair	Correlation Coefficient (linear fit)	
		Plot A	Plot B
blue •	Height (PHE) / Height (AMI)	0.006	0.281
black •	Height (PHE) / MCR1 (Bac.)	0.166	0.276
green •	Height (PHE) / MCR2 (Bac.)	0.033	0.349
red •	Height (AMI) / MCR1 (Bac.)	0.818	0.360
cyan •	Height (AMI) / MCR2 (Bac.)	0.968	0.810
orange •	MCR1 (Bac.) / MCR2 (Bac.)	0.894	0.419

Figure 7-6 Four candidate 'reference' measurements for normalisation of Raman spectral data for determination of polyphosphate. These comprise the peak heights of characteristic phenylalanine (PHE) and amide I (AMI) bands (at ~1002 and ~1657 cm^{-1} respectively), and scores on the bacterial-type component for both MCR models. Paired correlation plots are shown for direct comparison between the methods. Correlation plot A compares the mean of each sample (representing 50 spectra) and correlation plot B compares the ratio of the mean to the range (5^{th} to 95^{th} percentile) for each sample. Further explanation of the plots are given in the text.



Symbol	Correlation Pair	Correlation Coefficient (linear fit)	
		Plot A	Plot B
blue •	Height (PP2/PHE) / Height (PP2/AMI)	0.820	0.390
black •	Height (PP2/PHE) / MCR1 (PP/Bac.)	0.180	0.121
green •	Height (PP2/PHE) / MCR2 (PP/Bac.)	0.727	0.452
red •	Height (PP2/AMI) / MCR1 (PP/Bac.)	0.504	0.002
cyan •	Height (PP2/AMI) / MCR2 (PP/Bac.)	0.960	0.932
orange •	MCR1 (PP/Bac.) / MCR2 (PP/Bac.)	0.506	0.008

Figure 7-7 Four candidate normalised measurements for the determination of polyphosphate using Raman spectra. These comprise the ratios of the 'direct' measurements presented in Figure 7-5 to the 'reference' measurements in Figure 7-6. Paired correlation plots are shown for direct comparison between the methods. Correlation plot A compares the mean of each sample (representing 50 spectra) and correlation plot B compares the ratio of the mean to the range (5^{th} to 95^{th} percentile) for each sample. Further explanation of the plots are given in the text.

The ‘reference’ measurements in Figure 7-6 all represent Raman bands arising from bacterial constituents (phenylalanine, amides, and whole bacteria fingerprints). The box plots and correlation data indicate that some of these measurements differ from each other substantially.

Peak height of phenylalanine (PHE) stands out as a measurement that is poorly correlated with the other three measurements. This is shown by the fact that the correlation coefficients for phenylalanine with the other methods in both correlation plot A and correlation plot B range from just 0.006 to 0.349. On closer inspection, the box plot for phenylalanine shows a relatively large range in intensity and three extreme outliers in the spectral data sampled at time points 90, 150 and 240 minutes. The peak height data alone cannot explain these anomalies. Analysis of a plot of the residuals from MCR2 (Figure 7-8), however, reveals spectral components that are not fit by the model. These indicate the presence of interferents in the Raman spectra at around ~1000 cm^{-1} i.e. in the same position as the phenylalanine band. These interferents are relatively intense in the samples from 90, 150 and 240 minutes, corresponding with the samples containing extreme outliers in the box plot. Thus, it is likely that the peak height measured at ~1002 cm^{-1} does not always represent solely phenylalanine, introducing potentially large errors into this measurement.

The MCR1 bacterial components (MCR1 (Bac.)) show good correlation with peak height of amide (AMI) and MCR2 bacterial components (MCR2 (Bac.)) with respect to the mean values (correlation plot A, coefficients 0.818 and 0.894 respectively). The similarity in trends of the mean values for these three data sets can be seen in the box plots. With respect to the mean divided by the range, however, there is a much poorer correlation (correlation plot B, coefficients 0.360 and 0.419). The box plots show that some of the values for MCR2 (Bac.) are equal or very close to zero. The reason for this lies in the model loadings. The PP-type component for MCR1 contained some small bacterial contributions. This meant that in some Raman spectra, the bacterial contributions were small enough that they could be modelled sufficiently with the PP-type component alone and so a value of zero was assigned on the bacterial-type component.

The reference measurements made of AMI and MCR2 (Bac.) show good correlation with each other with respect to both mean values and the mean divided by the range (0.968 and 0.810 respectively). This similarity can be seen in box plots and furthermore, no extreme outliers (such as those in the phenylalanine measurement) are evident.

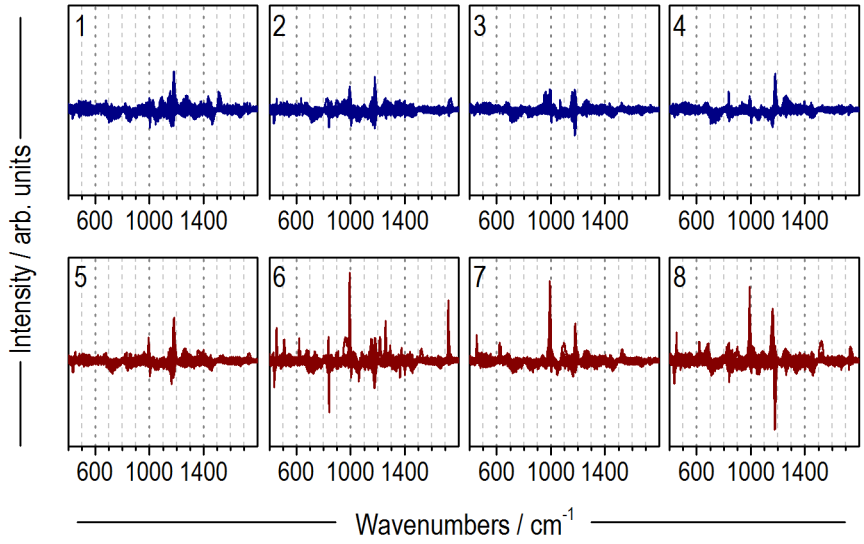


Figure 7-8 Residuals from MCR model 2 (MCR2) applied to Raman spectra of samples from SBR-D during the uptake release test described. Samples numbered 1 to 8 correspond sequentially to different time points in the test: 0, 7, 18, 45, 60, 90, 150 and 240 minutes. For each sample, the residuals from 50 Raman spectra are overlaid. Blue and red plots indicate samples taken from the anaerobic and aerobic phase respectively.

The normalised measurements representing 'direct' measurements divided by reference measurements are shown in Figure 7-5. As expected from the previous discussion, using PHE and MCR1 (Bac.) as reference measurements give rise to some unusual results and as such are poor candidates for a normalisation method. In particular, the box plots show that normalised MCR1 (PP / Bac.) data has some large outliers. At the 240 minute time point, for example, one data point is almost 30 times larger than the mean. These extreme outliers are caused by the fact that some of the scores on the MCR1 (Bac.) reference tended to zero. In contrast, the data normalised by AMI and MCR2 (Bac.) appeared to be similarly successful and showed similar trends in and between samples (correlation coefficients 0.960 and 0.932 respectively). This suggests that both these approaches are good candidates for normalising the measurement of polyphosphate with Raman spectroscopy.

The summed data from each set of 50 samples is presented in Figure 7-9 for four types of analysis. The 'direct measurements made with peak height data (PP2) and with MCR2 (PP-type component) can be seen to be very similar to each other. Likewise the two normalised measurements made by the ratio of peak heights for PP and AMI, and the ratio of scores for

MCR2 PP- and bacterial-type components can be seen to be very similar to each other. The normalised data is similar to the absolute data except for the first two sample points where the normalised data is lower than the absolute data. Since the data was acquired in one 'batch run' on the same day, the direct measurements may be a reasonable representation of the true variation in polyphosphate in the samples analysed. The similarity between the absolute measurements and normalised measurements is thus encouraging.

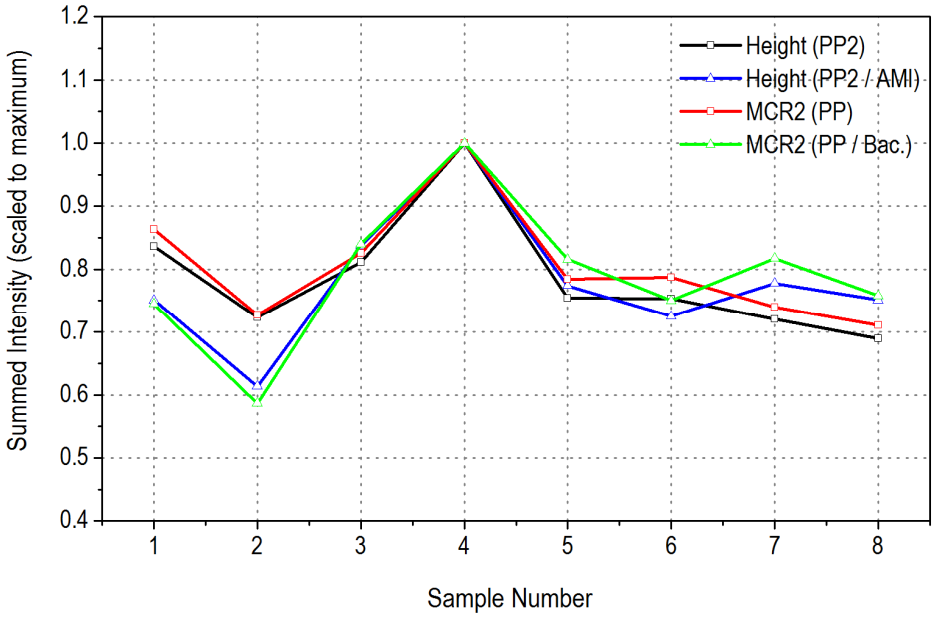


Figure 7-9 Summed polyphosphate quantity measured by Raman spectroscopy of 50 individual cells for each sample taken during the uptake-release test. The results for four different methods of data analysis are shown. For ease of comparison the data is scaled so that the maximum value for each set of results is 1. Samples numbered 1 to 8 correspond sequentially to different time points in the test: 0, 7, 18, 45, 60, 90, 150 and 240 minutes.

Overall, it has been argued that normalisation is a necessary step in spectral data analysis for the quantification of polyphosphate with Raman spectroscopy. Of the four approaches to normalisation investigated here, one peak height method (referencing to the amide I band) and one multivariate approach based on MCR (referencing to the modelled bacterial-type component) were found to perform similarly well. Of these, the multivariate approach may offer some advantages such as being able to analyse model residuals to identify unusual Raman spectra (and hence potentially erroneous results). Furthermore, the amide I band

represents just one constituent of bacteria and the relative concentration of this between bacterial cells may change. On the other hand, the MCR model uses something more like an overall representation of all the bacterial constituents and hence may prove a more consistent reference point for polyphosphate quantification.

Variation in Intracellular Polyphosphate during EBPR Cycle

The uptake-release test from which samples were taken for Raman spectroscopic analysis showed characteristic EBPR behaviour in terms of the concentrations of COD and orthophosphate measured in the mixed liquor. It could be expected, therefore, that the quantity of polyphosphate measured within EBPR bacteria should also reflect the expected trend – that cellular polyphosphate should decrease during the anaerobic phase and increase during the aerobic phase. This trend was observed by Majed *et al.* who also used Raman spectroscopy to measure the changing cellular polyphosphate concentrations in EBPR bacteria during an uptake-release test [68, 72]. Here, however, the expected trends were not observed – either in individual bacteria or in the summed data. One reason for this might be the issue of subsampling as shown in Chapter 4. The measured laser spot size of the Raman instrument used in this work was approximately 1.2 µm (in the lateral direction) and hence a little smaller than most of the cells analysed in this work. Further investigation of the variability of polyphosphate quantity measured within cells using the technique described here is required before meaningful conclusions can be made about the variability of polyphosphate quantity between samples. Due to the heterogeneity of bacterial cells, one useful parameter to measure that has a degree of consistency, would be the quantity of polyphosphate per cell. This would require either the use of Raman mapping or an instrumental set up that provides a slightly larger laser spot size. A larger laser spot size can be achieved by reducing the numerical aperture of the objective used. This also increases the depth of field of the laser focal spot, however, so could decrease signal-to-background ratio if more of the substrate is illuminated relative to the sample.

Another reason for not observing the expected trends in cellular polyphosphate might be the difficulty of measuring a representative cross-section of the EBPR bacterial population. There is bias in the particular cells selected for analysis purely by their arrangement on the CaF_2 substrate. Nicely presented cells situated apart from other cells are ideal candidates for analysis. Furthermore with the degree of heterogeneity demonstrated in the bacterial population, it is suspected that a much greater sample size than 50 cells is required to

adequately represent the polyphosphate content in the PAO population. This may be achieved by the development of high-throughput analysis for single cells, or alternatively, by switching approach to analysing many cells per sample (such as in a cell pellet) rather than individual cells.

7.4 ESTIMATION OF LOD AND LOI OF POLYPHOSPHATE

7.4.1 LOD and LOI of Aqueous Polyphosphate Solutions

Raman spectra obtained for aqueous polyphosphate solutions of varying concentration are shown in Figure 7-10. Furthermore, two sets of spectra are shown which correspond to those acquired with 10 and 200 second acquisition times respectively. On each plot, both the raw spectrum and the background/baseline-corrected spectrum is shown. The raw spectra contain two broad Raman bands at ~1372 and 1633 cm^{-1} that are assigned to the aqueous matrix background contribution since they are also present in the pure water samples. It can be seen in the corrected spectra that these bands have been removed effectively by the background and baseline preprocessing steps.

From the spectra presented, two things are immediately apparent: First, that the polyphosphate spectrum with characteristic bands at ~690 and ~1150 cm^{-1} are clearly visible at the highest concentration (0.2 M), and that these bands decrease in intensity as the concentration is lowered. Second, the spectra obtained with shorter acquisition times (10 seconds) have lower signal-to-noise ratio than the spectra obtained with a longer acquisition time (200 seconds). Thus, the polyphosphate concentration LOD decreases as acquisition time increases.

If LOD is defined by the concentration at which the most intense polyphosphate peak (at ~1150 cm^{-1}) is distinguishable from noise, then this is observed in the spectra presented to be approximately 0.05 M for a 10 second scan and less than 0.05 M for a 200 second scan. For more complex spectra it may be more beneficial to define another limit – the LOI – where both characteristic polyphosphate peaks are distinguishable from noise. This can be seen to lie somewhere between 0.1 and 0.2 M for a 10 second scan, and at approximately 0.05 M for a 200 second scan. From the peak height ratio of the two polyphosphate bands, it might be expected that LOI is roughly 2 to 3 times the LOD. Further work would be required to confirm this. In addition, in this preliminary study, 0.05 M was the lowest concentration of polyphosphate measured, but in future work, lower concentrations could and should be

investigated. To simplify our discussion, just LOD (and not LOI) is considered in the following section.

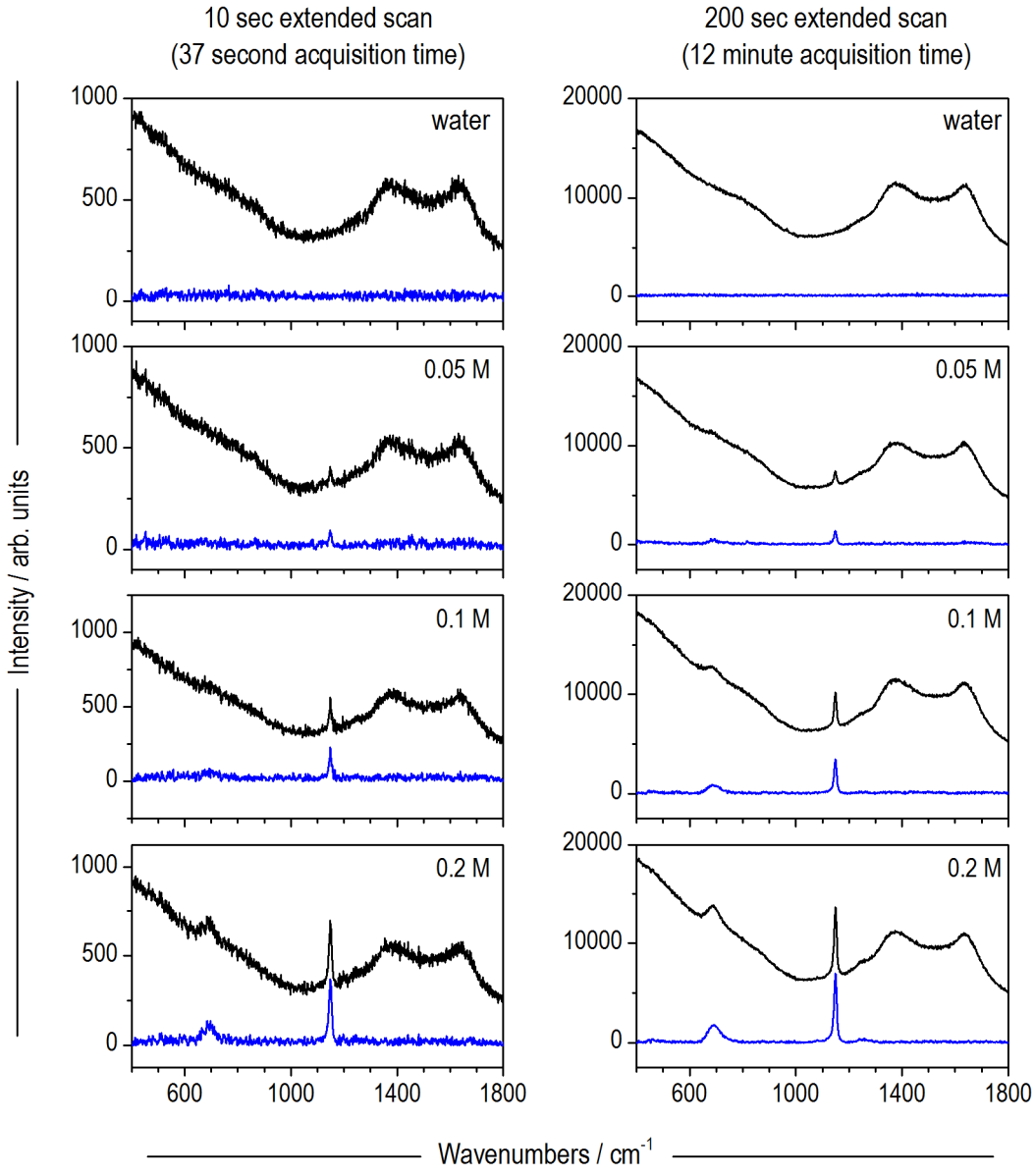


Figure 7-10 Raman spectra of aqueous polyphosphate solutions of varying concentration taken with two different acquisition times. Both raw spectra (in black) and background-baseline corrected spectra (in blue) are shown.

For this initial approximation, LOD was calculated according to a definition given in the International Conference on Harmonization Q2 (ICH Q2) Guidelines [211], where LOD is equated simply to twice the mean of the measured blank value. In this case, the blank used was ultrapure (Milli-Q) water so that LOD was calculated thus:

$$LOD = 2 \times mean\ signal\ of\ pure\ water \qquad \textit{(Equation 7-1)}$$

Raman spectra were acquired from four ultrapure water samples measured on different days. From these, LOD was calculated using both peak height data and (for comparison) PLS model data. For measurement by peak height, the signal intensities at ~697 and ~1172 cm^{-1} (i.e. corresponding to both polyphosphate peak positions) were averaged for each measurement so that effectively the blank measurement was derived from four samples but eight data points. For PLS model data the mean water signal was defined as the mean of the scores assigned to the four blank samples in the model built as described in Section 7.2.5. These mean signals were converted to concentration values based on the regression analysis of the peak height data and PLS model respectively. Estimation of LOD based on visual inspection, peak height data and PLS model data are summarised in Table 7-2. It can be seen that the results agree well with each other and from these, taking the most conservative estimation, the LOD was approximated to be 0.08 M for a 10 second scan and 0.02 M for a 200 second scan.

Table 7-2 LOD calculated for detection of aqueous polyphosphate with the Raman instrument, set-up and sample presentation described in this study.

LOD Concentration / M					
Peak Height Data		PLS model data		Visual Inspection	
10 sec	200 sec	10 sec	200 sec	10 sec	200 sec
0.05	0.01	0.08	0.02	0.05	< 0.05

7.4.2 LOD of Polyphosphate in EBPR Bacteria

The LOD estimated based on measurements of aqueous polyphosphate solutions cannot be directly equated to an LOD for the detection of polyphosphate in EBPR bacteria. Firstly, as discussed in Chapter 6, the Raman signal of polyphosphate has a significant matrix effect, and the two Raman bands in bacterial samples were found to be shifted approximately ten wavenumbers with respect to those in pure aqueous polyphosphates. Secondly, the presentation of the aqueous samples for analysis was different to that of bacterial samples. This is depicted in Figure 7-11 and discussed below.

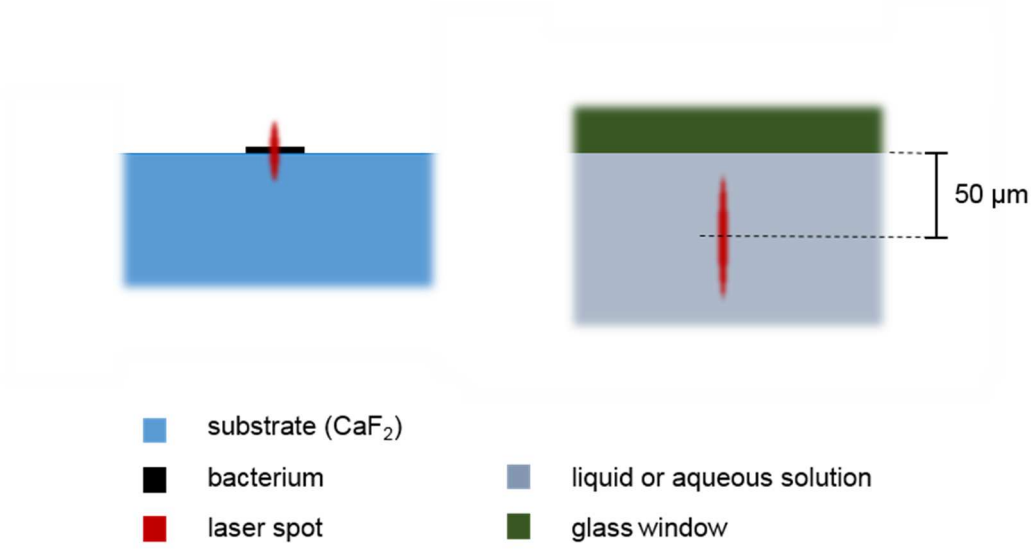


Figure 7-11 Schematic diagram (not to scale) of a bacterium sample on a CaF_2 substrate (left) and a liquid or aqueous sample enclosed in a sample well with a glass coverslip (right). The optical path travelled by the laser is different in each case – onto the surface of the CaF_2 substrate to measure a bacterium, but through an MgF_2 window to measure an aqueous sample. As such the laser spot size can be expected to vary between sample types.

Bacterial samples were dried on a CaF_2 substrate and the laser was focussed directly onto the surface of the sample. The laser spot dimensions for this set-up were determined experimentally (in Chapter 4) to be 1.2 µm in the lateral plane and 4 µm in the axial plane, yielding an estimated laser sampling volume of 4.5 μm^3. The bacteria did not fill all the available sampling volume in the axial plane. For example, based on a 50 µm × 50 µm AFM image (Figure 7-12), an estimation of the height of dried EBPR bacteria on a glass slide was

made. In three different regions of the image, line profiles of height data were acquired. In addition to this, at selected points on these profiles, minimum (corresponding to substrate only) and maximum (corresponding to bacterial presence) height data was reported. Based on this data (shown in Figure 7-12), 0.5 μm was considered to be a conservative estimate of the height of a dried EBPR bacterium on a substrate. Since the EBPR bacteria we observed by microscope in the course of this work were typically 2 to 5 μm in diameter, 0.5 μm is a reasonable height estimate once dehydration is taken into account. This sample height corresponds to (and hence fills) just one eighth of the measured laser spot size in the axial plane.

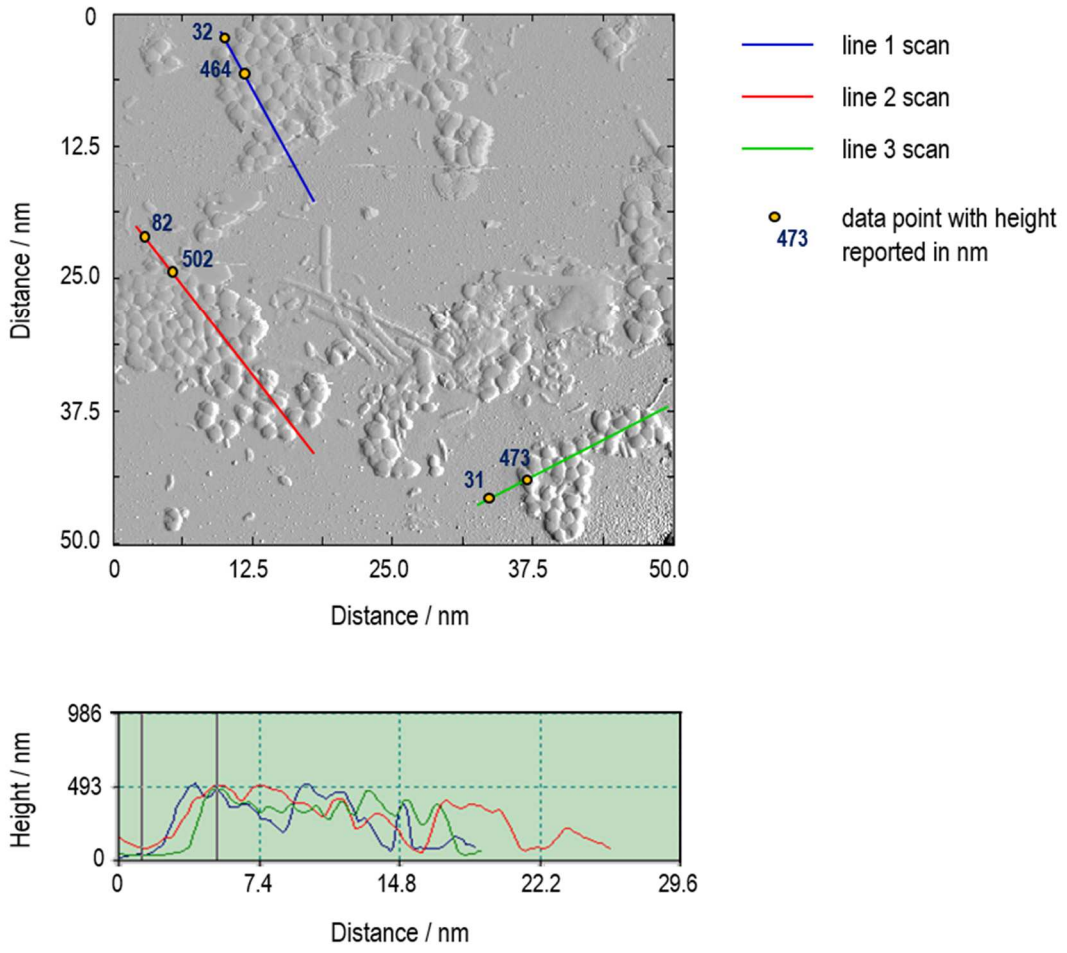


Figure 7-12 AFM image of EBPR bacteria dried on a glass microscope slide (top). Height data recorded from the three different line profiles indicated on the AFM image (bottom).

Aqueous samples were placed in a small slide well and sealed with a glass window to prevent evaporation of the sample during acquisition. The laser spot dimensions for this set-up were determined experimentally to be 1.2 µm in the lateral plane and 11 µm in the axial plane, yielding an estimated laser sampling volume of 12.5 μm^3. The laser spot was focussed 50 µm into the sample from the underside of the glass window and the aqueous sample filled the entire sampling volume.

It can thus be seen that inferring a LOD for polyphosphate in bacteria from a measurement made on pure aqueous polyphosphate solutions is difficult, if not impossible. However, as a starting point, an approximation of LOD for this method has been provided here and is expressed below (Table 7-3) in mass units and per μm^3 rather than per Litre to be more appropriate for the scale of the Raman sampling volume and polyphosphate content in individual bacteria.

Table 7-3 LOD of aqueous polyphosphate measured with Raman spectroscopy for two different acquisition times and expressed in different concentration units.

LOD Unit	10 sec acquisition	200 sec acquisition
mol (PO_4) / L	0.08	0.02
pmol (PO_4) / μm^3	8×10^{-5}	2×10^{-5}
pg (P) / μm^3	25×10^{-4}	6×10^{-4}

One step toward improving the LOD measurement would be to measure well-defined pure polyphosphate samples with the same sample presentation as the bacterial samples. One method to achieve this may be to deposit droplets of polyphosphate solution of known concentration onto a slide. The droplet would then contain a defined mass of polyphosphate. Since the sample prepared in this way would be heterogeneous it would require a Raman mapping measurement to measure the entire sample. This approach was used by Konorov *et al.* for the determination of glycogen content in cells using Raman spectroscopy [162]. This method has its own difficulties, however, such as how can the intensities measured in a mapping measurement be related to a single point measurement, and how can (ideally) a micro-size droplet be deposited small enough that it can be mapped adequately and with

little introduction of error arising from factors such the microscope stage not being perfectly flat.

7.5 SUMMARY AND CONCLUSIONS

Raman spectroscopy has some significant advantages for application to the quantification of polyphosphate in bacteria. Despite this, it has not yet been used extensively for this purpose. In cases where this method has been used, analysis methods are either qualitative or relate polyphosphate quantity directly to peak height. Due to the variability inherent in Raman spectral measurements, however, some kind of spectral normalisation is vital in order to render Raman-based quantification of polyphosphate more universally comparable. In this study we compared for the first time four options for normalising the Raman spectra of EBPR bacteria for the purpose of quantifying polyphosphate: two methods based on peak height ratios and two novel methods based on MCR analysis. From these, one peak height method (referencing to the amide I band) and one multivariate approach based on MCR (referencing to the modelled bacterial-type component) were found to perform similarly well. Between these two, the multivariate approach is recommended since it offers the advantage of being able to analyse model residuals to identify unusual samples, and furthermore, the reference signal comprises an overall representation of the bacterial constituents rather than a single bacterial component that itself could change in concentration between samples. Finally, to aid data interpretation and also the comparison of the Raman-based quantification method with other techniques, an estimation of LOD was made. For the Raman instrument used in this work, the LOD of aqueous polyphosphate samples was estimated to be 0.08 M and 0.02 M for a 10 second and 200 second extended scan respectively. The problems encountered in inferring an LOD for polyphosphate in bacteria from such a measurement were discussed. In the absence of better methods for determining LOD in these systems at the current time, however, these values serve as a benchmark figure to which further measurements may be compared.

Chapter 8
Summary and Conclusions

The overarching aim of this thesis was to investigate the potential application of Raman spectroscopy to further understand and improve EBPR. More specifically, the research questions addressed in this work were:

1. What is an effective, automated method for preprocessing whole Raman spectra of EBPR bacteria in preparation for exploratory analysis using multivariate statistics?
2. What information may be gained from qualitative analysis of Raman spectra of EBPR bacteria that could yield new insights to the process or open up new avenues of research?
3. How can the Raman spectra of EBPR bacteria be analysed to yield quantitative data?

These questions were addressed in Chapters 5, 6 and 7 respectively:

In Chapter five, the necessity and challenge of pre-processing Raman spectra of EBPR bacteria was discussed. An overview and framework for categorising existing baseline correction methods was presented to aid navigation of the (sometimes confusing) literature on this subject. Following this, a comparison study of nine different baselining methods was conducted on a sample set of Raman spectra. Performance of the different methods was compared with respect to visual appearance of the spectra and how they may affect subsequent qualitative and quantitative analysis. This was the first study of its type, investigating preprocessing of EBPR bacteria Raman spectra in detail. Furthermore it is one of very few studies comparing the application of a number of different baselining techniques to real data rather than synthetic data.

The results of the comparison study indicated firstly that Raman spectra of EBPR bacteria that were corrected with both background and baseline removal performed better than background-only-corrected or uncorrected (‘raw’) data with respect to appearance and qualitative and quantitative analysis. Of the variety of baselining methods trialled, ‘rolling ball’, ‘local medians’, ‘Beier’ and ‘robust baseline estimate’ did not appear to fit a sample spectrum well whatever model parameters were employed. These techniques were thus not investigated in detail in this study. Of the remaining five baselining methods trialled, none seemed to outperform the others consistently or significantly in the qualitative and quantitative models applied to various data sets. Visually, the methods based on spectra-wide polynomial fitting (‘modified polynomial’ (ModPoly) and ‘adaptive minmax’ (AMM)) were shown to be prone to inconsistencies in how tightly they fit adjacent regions of the spectra due to the characteristics of this type of function. On the other hand, ‘small-window moving

average' (SWiMA) and 'alternating least squares' (ALS) appeared to fit adjacent regions of the spectra with similar tightness. SWiMA gave the least smooth baseline but (together with AMM) fit the spectra best in extreme regions. In contrast, the baselines calculated by ModPoly, ALS and 'adaptive-weight penalised least squares' (APLS) deviated from the spectra at the extremes in one or more cases. Since the fit at the extremes is important for the spectra investigated here, SWiMA was selected as the baselining technique to use for the remainder of this study.

In future work, for the reasons of performance, smoothness and calculation time, ALS may the preferred technique to use if the fit at the spectral extremes is either not so critical or could be improved. Another approach which may be worth investigating could be to use the SWiMA algorithm to identify zero-signal regions in each spectrum and to connect these regions with a low-order polynomial. Finally, for full-automation of these preprocessing steps, methods for automatic removal of cosmic rays from spectra should be investigated in future work to replace manual removal of these spectral artifacts.

In Chapter six a qualitative examination was conducted of the Raman spectra acquired from a large number of individual bacteria sampled from three different EBPR reactors. In line with a handful of previous studies, the capability of Raman spectroscopy to identify three metabolites key to the EBPR process – polyphosphate (PP), polyhydroxyalkanoates (PHA) and glycogen – in these bacteria was confirmed. Following this, the variability of the characteristic Raman signals for these compounds between samples was investigated in detail for the first time. The position of the major Raman bands for glycogen and PHA were not seen to vary much in spectra of the pure compounds versus spectra of these metabolites in EBPR bacteria. Furthermore, these peak positions were not seen to vary much between the bacterial samples themselves. In contrast, the position of the major polyphosphate Raman bands were shown to vary significantly between pure compound samples and bacterial samples as well as between the bacterial samples themselves. This could raise some challenges for the analysis of EBPR bacteria by Raman spectroscopy. For example, if quantitative analysis of polyphosphate content were desirable, what kind of compound could be used as a standard when a pure polyphosphate sample has such a different signal to that arising from intracellular bacterial polyphosphate? On the other hand, perhaps these peak position shifts could be a useful indicator of some parameter in the cells themselves. This is something that could be considered in future work.

The Raman spectra of EBPR bacteria are complex with many overlapping signals. Despite this, multivariate statistical tools have not yet been applied to these kind of samples. Here, for the first time, two multivariate analytical tools – multiplicative curve resolution (MCR) and partial least squares-discriminant analysis (PLS-DA) – were applied to sample sets of Raman spectra of individual EBPR bacterial cells for the first time.

Using MCR, over 95% of the variance in the Raman spectra could be modelled successfully with four major components. These four components closely resembled pure spectra of the three key EBPR metabolites – polyphosphate, PHA and glycogen – plus a spectrum that could reasonably be considered to represent the underlying bacterial 'matrix'. For each Raman spectrum (of an individual EBPR bacteria) analysed, a 'score' was assigned by the model for each of these components. Plotting these scores gave an overall visual impression of the combination of metabolites present in each sample and the distribution of these among a data set. This could be a useful tool for making quick comparisons of metabolite content in individual cells for large datasets. Furthermore, it would be interesting to compare these plots for samples analysed under different operating conditions. Perhaps they would constitute a kind of 'fingerprint' that is characteristic of a particular reactor or operating conditions and may then further be related to process performance. The success shown by MCR in extracting components so similar to the pure compound spectra and the underlying bacterial matrix was something that was also considered to be potentially applicable to quantitative analysis of the spectra. This was investigated further in Chapter 7.

Using PLS-DA, the automated identification of candidate PAOs based on their Raman spectra was investigated for the first time. This kind of analysis could be useful in an application such as high-throughput Raman activated cell sorting – to separate PAOs from the mixed EBPR microbial community so that they can be studied in isolation. For the purpose of the analysis, any bacterium giving rise to a spectrum containing a polyphosphate signal was considered to be a candidate PAO. Compared to a similar analysis made using peak height data, PLS-DA was found to be less sensitive to variation in polyphosphate peak position and misclassified very few spectra. The option to analyse residuals makes PLS-DA an attractive method with respect to being able to identify samples that are unusual and which should potentially be rejected on that basis. For this kind of application, therefore, PLS-DA is preferred over analysis based on peak height.

In Chapter seven the application of Raman spectroscopy to the quantification of polyphosphate in individual EBPR bacteria was investigated. In the few cases where this method has previously been used, the spectral analysis has typically related polyphosphate quantity directly to peak height. Due to the variability inherent in Raman spectral measurements, however, some kind of spectral normalisation is vital in order to render Raman-based quantification of polyphosphate more universally comparable. In this study we compared for the first time four options for normalising the Raman spectra of EBPR bacteria for the purpose of quantifying polyphosphate: two methods based on peak height ratios and two novel methods based on MCR analysis. To the best of the author's knowledge, MCR has not previously been applied to a normalisation problem such as this. From amongst the four normalisation methods investigated, one peak height method (referencing to the amide I band) and one multivariate approach based on MCR (referencing to the modelled bacterial-type component) were found to perform similarly well. Between these two, the multivariate approach is recommended since it offers the advantage of being able to analyse model residuals to identify unusual samples, and furthermore, the reference signal comprises an overall representation of the bacterial constituents rather than a single bacterial component that itself could change in concentration between samples.

Finally, to aid data interpretation and also the comparison of the Raman-based quantification method with other techniques, an experimental estimation was made of the limit of detection of polyphosphate using Raman spectroscopy. With the instrumental set-up in this work, the LOD of aqueous polyphosphate samples was estimated to be approximately 0.08 M and 0.02 M for a 10 second and 200 second extended scan respectively. The problems encountered in inferring an LOD for polyphosphate in bacteria from such a measurement was discussed. In the absence of better methods for determining LOD in these systems at the current time, however, these values serve as a benchmark figure to which further measurements may be compared. Future work should focus on researching ways in which a more comparable form of polyphosphate 'standard' might be prepared so that direct correlation can be drawn between measurements made on this standard and measurements made in bacterial cells. Furthermore, perhaps a 'standard relationship' could be defined between measurements made on such a standard and measurements made on a more easily available material so that LOD could readily be determined on any Raman system. In addition to estimation of LOD, this would also offer a way forward for developing a fully-quantitative measurement of polyphosphate in bacteria, rather than the semi-quantitative measurement described in this work.

In response to the research questions posed at the start of this work, the main conclusions are summarised below:

1. Background and baseline removal are necessary preprocessing steps to enable subsequent analysis of the Raman spectra of individual EBPR bacteria to be comparable with each other. Of the nine baselining methods studied, five were found to perform similarly well but SWiMA was determined to be the best technique at the current time for baselining these spectra.

2. The signatures of the three key EBPR metabolites – polyphosphate, PHA and glycogen are clearly identifiable in individual EBPR bacteria when present. In contrast to the two other metabolites, the position of characteristic bands of polyphosphate vary significantly between samples. This is one parameter which may be investigated in future as a function of EBPR operating conditions and performance. The Raman spectra of individual bacteria can be modelled with multivariate tools to open up new areas for research. For example, PLS-DA can automatically recognise candidate PAOs based on the presence of a polyphosphate signal in the Raman spectra. This could be applied in something like high-throughput Raman-activated cell sorting to separate live PAOs from the mixed EBPR microbial community so that they can be studied in isolation.

3. Spectral normalisation is necessary to allow comparison of Raman-based quantification of polyphosphate in bacterial cells. MCR modelling offers a novel way to normalize the Raman spectra of individual EBPR bacteria and has the added benefit of being able to analyse model residuals to identify unusual samples. With the instrumental set-up in this work, the LOD of aqueous polyphosphate samples was estimated to be approximately 0.08 M and 0.02 M for a 10 second and 200 second extended scan respectively. Future work is required to research ways in which a more comparable form of polyphosphate ‘standard’ might be prepared so that direct correlation can be drawn between measurements made on this standard and measurements made in bacterial cells. This would not only provide a way to measure LOD for the method, but would also pave the way ahead for developing a fully-quantitative measurement of polyphosphate in bacteria, rather than the semi-quantitative measurement described in this work.

Overall, several applications and challenges of Raman spectroscopy for the investigation of EBPR bacteria have been presented in this work together with recommendation for how to process the spectral data. An obvious next step would be to attempt to apply these tools to analysing live bacteria in order to exploit the non-destructive nature of Raman spectroscopy and investigate how it might be useful for in-line EBPR process monitoring. The conclusions drawn from this work indicate that Raman spectroscopy could be used as a tool to improve EBPR but more work is required to refine and apply these methods.

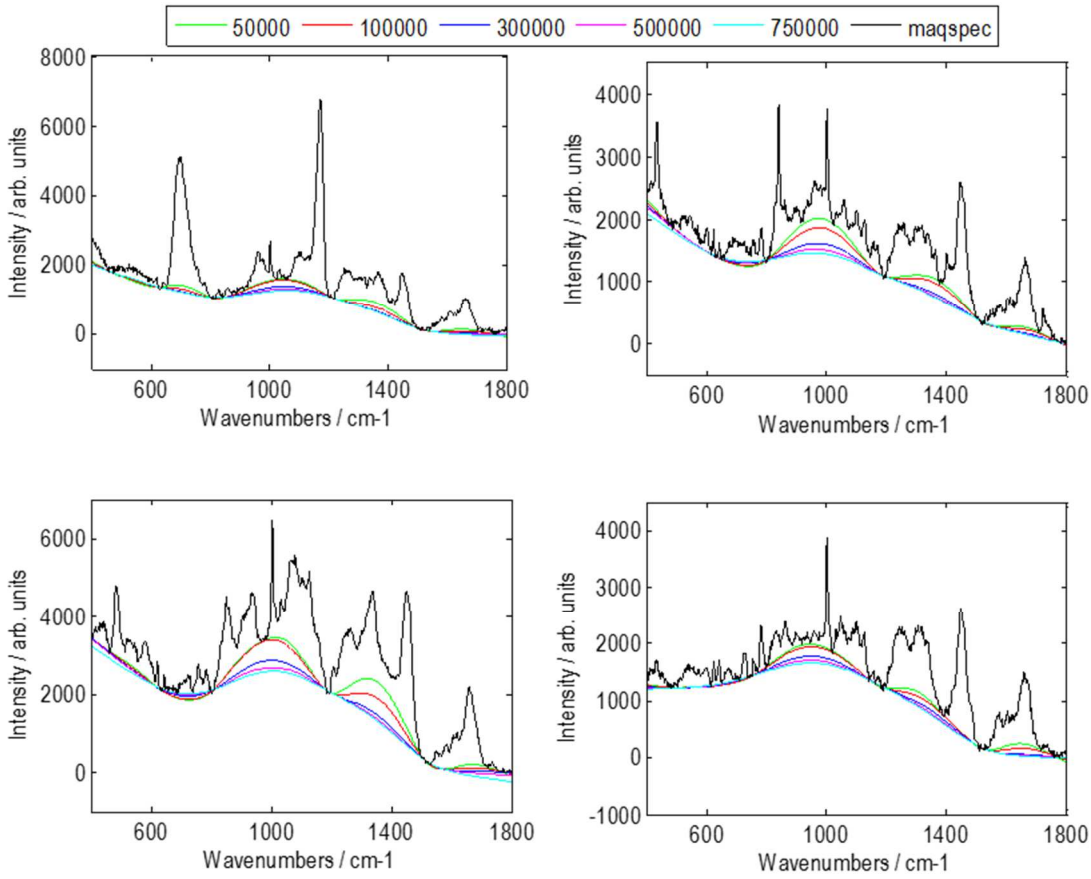
Appendix 1

SBR Rig Components

	Component	Manufacturer	Supplier	Code	Cost per unit excl. VAT	Qty
Pumps	Peristaltic pump drive L/S 24VDC	Masterflex	Cole-Parmer	OU-07533-70	352.00	2
	Pump head L/S Easy-Load II stainless	Masterflex	Cole-Parmer	EQ-77200-62	208.00	4
	Mounting hardware for 2 heads	Masterflex	Cole-Parmer	EW-77200-02	25.00	2
	Peristaltic tubing Pt-cured silicone #LS36 (3/8")	Masterflex	Cole-Parmer	EQ-96400-36	93.00	1
Stirrer	Propeller 2" diameter, 5/16" bore, 316SS		Cole-Parmer	EW-04552-50	33.00	2
	DC motor 24V	Canon	Trident Eng. Ltd.	3-38/1424	95.82	2
	6:1 gearhead	Trident	Trident Eng. Ltd.	GS38.0006	incl. with above	2
	propeller shaft, motor mounting and housing	in-house				
Fittings etc	3/8" 2-way stainless ball valve	HAM-LET	FTI Ltd.	H800L-SSL3/8	92.44	2
	3/8" 3-way stainless ball valve	HAM-LET	FTI Ltd.	H800L-SSL3/8-T	126.45	2
	3/8" stainless elbow union	HAM-LET	FTI Ltd.	765LSS3/8	13.86	4
	3/8" stainless male connector	HAM-LET	FTI Ltd.	768LSS3/8X3/8	7.78	4
	3/8" stainless tube, 3 metres	HAM-LET	FTI Ltd.	SST6-20	8.23	1
	semi-flexible tubing 10mm ID	RS	Stores			
	any other fittings	in-house				
Monitoring	DO sensor (3m cable)	Ponsel	Env. Mon. Solns.	PF-CAP-C-00140	615.60	2
& Control	pH sensor (3m cable)	Ponsel	Env. Mon. Solns.	PF-CAP-C-00143	409.50	2
	pH sensor cartridge	Ponsel	Env. Mon. Solns.	PF-CAP-C-00155	176.00	2
	k-type insulated thermocouple 3x250mm		RS	3971270	18.43	2
	miniature float switch		RS	519232	4.61	2
	housing+mounting for sensors	in-house				
	power supply unit	in-house				
	units to interface hardware to LabView	in-house				
	PC with labview	-	SEE support		0.00	1
Containers	30L Influent Container (stackable jerrycan)	Ampulla	Ampulla	S30LSC1Nat	8.03	2
& Infrastructure	60L Effluent Container ("open top keg")	Ampulla	Ampulla	60L1	18.66	2
	3.5L glass reactor container (droppar jar)	Ikea	Ikea	40112543	5.21	2
	Scissor stand	Sealey	Amazon	MC5908	66.63	1
	acrylic 10mm thick mounting plate	in-house				
	Fridges					2
Aeration	Aquarium air pump	Tetratec	Amazon	APS50	8.86	4
	Air stones		Amazon		3.38	10 pcs
	air in-let line sheath (stainless steel)	in-house				

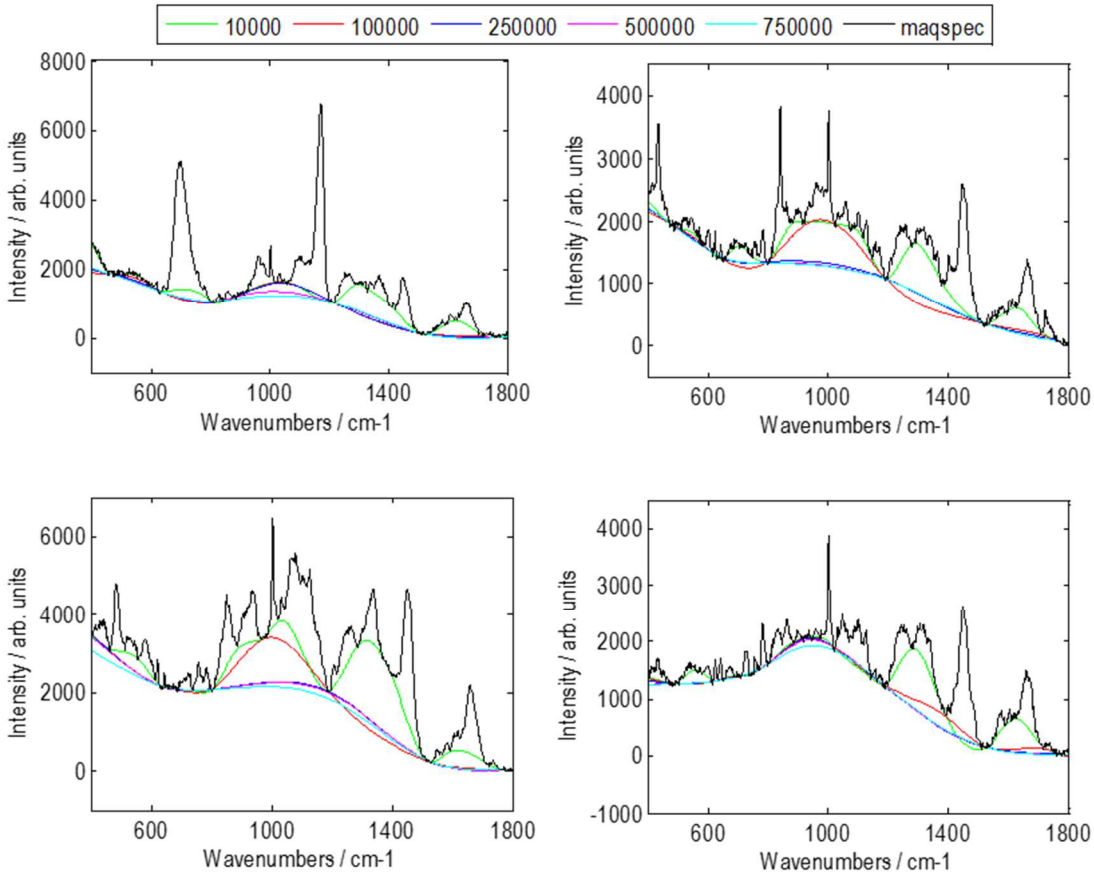
Appendix 2
Comparison of Parameters for Selected Baselining Methods

Asymmetric Least Squares



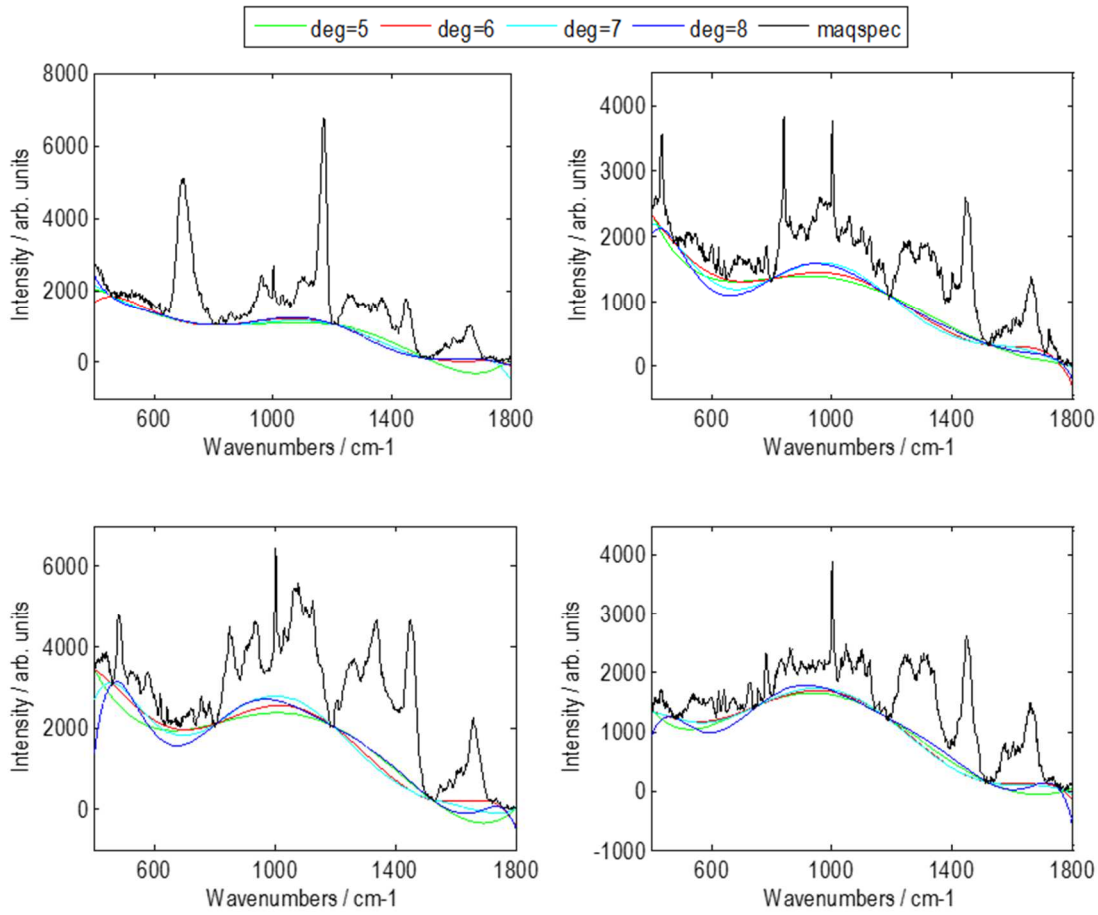
Comparison of smoothing parameter value used in baselining
with the asymmetric least squares method

Adaptive-Weight Penalised Least Squares



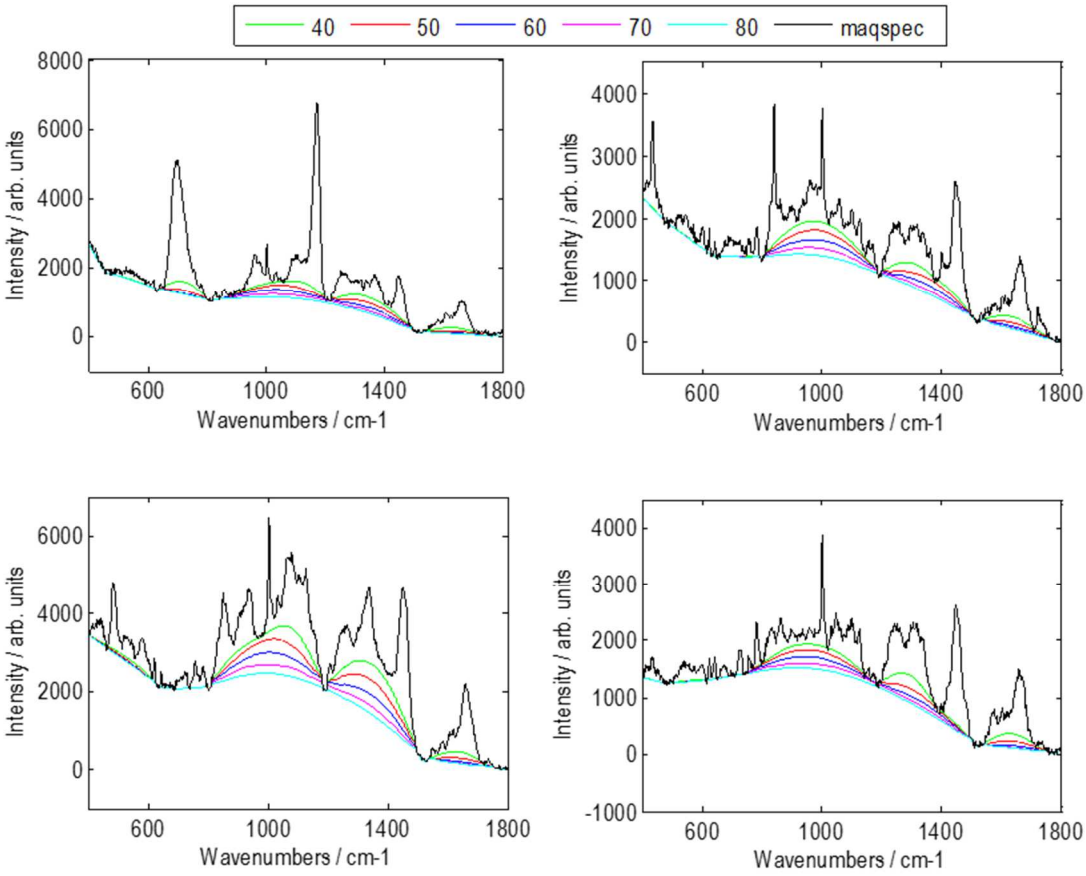
Comparison of smoothing parameter value used in baselining
with the adaptive-weight penalised least squares method

Modified Polynomial



Comparison of polynomial degree value used in baselining
with the modified polynomial method

Small-Window Moving Average



Comparison of number of iterations used in baselining
with the small-window moving average method

References

1. Gilbert, N. The Disappearing Nutrient. *Nature* **461**, 716-718 (2009).
2. Cordell, D., Drangert, J.-O., and White, S. The Story of Phosphorus: Global Food Security and Food for Thought. *Global Environmental Change-Human and Policy Dimensions* **19**, 292-305 (2009).
3. Van Kauwenbergh, S.J., *World Phosphate Rock Reserves and Resources. Technical Bulletin IFDC T-75.* International Fertilizer Development Center, Muscle Shoals, AL,
4. Edixhoven, J.D., Gupta, J., and Savenije, H.H.G. Recent Revisions of Phosphate Rock Reserves and Resources: A Critique. *Earth System Dynamics* **5**, 491-507 (2014).
5. Jasinski, S., *U.S. Geological Survey Mineral Commodity Summaries: Phosphate Rock.* 2015.
6. Manning, D.A.C. How Will Minerals Feed the World in 2050? *Proceedings of the Geologists Association* **126**, 14-17 (2015).
7. Elser, J. and Bennett, E. A Broken Biogeochemical Cycle. *Nature* **478**, 29-31 (2011).
8. Ulrich, A.E. and Schnug, E. The Modern Phosphorus Sustainability Movement: A Profiling Experiment. *Sustainability* **5**, 4523-4545 (2013).
9. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the Review of the List of Critical Raw Materials for the EU and the Implementation of the Raw Materials Initiative. (2014).
10. Tchobanoglous, G., Burton, F.L., Stensel, H.D., & Metcalf, and Eddy, *Wastewater Engineering: Treatment and Reuse*. 2003, Boston: McGraw-Hill.
11. Seviour, R.J., Mino, T., and Onuki, M. The Microbiology of Biological Phosphorus Removal in Activated Sludge Systems. *FEMS Microbiology Reviews* **27**, 99-127 (2003).
12. Lund, J.W.G. Phosphorus and Eutrophication Problem. *Nature* **249**, 797-797 (1974).
13. Yoshida, H., Van Dijk, K.C., Drizo, A., Van Ginkel, S.W., Matsubae, K., and Buehrer, M., *Phosphorus Recovery and Reuse*, in *Phosphorus, Food, and Our Future*, K.A Wyant, J.R. Corman, and J.J. Elser, Editors. 2013, OUP: New York, NY.
14. Le Corre, K.S., Valsami-Jones, E., Hobbs, P., and Parsons, S.A. Phosphorus Recovery from Wastewater by Struvite Crystallization: A Review. *Critical Reviews in Environmental Science and Technology* **39**, 433-477 (2009).

15. Thames Water (28 September 2010) *Renewable Phosphate Fertiliser from Sewage: A UK First*. [Accessed: February 2016]; Available from: http://www.thameswater.co.uk/media/press-releases/10953.htm.

16. Thames Water (6 November 2013) *Fertiliser from Sewage Saving the Planet - and Cash*. [Accessed: February 2016]; Available from: http://www.thameswater.co.uk/media/press-releases/17393.htm.

17. Zakrzewski, M. Crystal Green, Environmentally-Responsible Phosphorus Fertilizer. *AgLink Magazine* **July/August 2012**, 6-7.

18. Jaffer, Y., Clark, T.A., Pearce, P., and Parsons, S.A. Potential Phosphorus Recovery by Struvite Formation. *Water Research* **36**, 1834-1842 (2002).

19. Ostara, *Case Study: Slough Sewage Treatment Works, Thames Water*. Ostara Nutrient Recovery Technologies Inc, http://beta.ostara.com/sites/default/files/pdfs/Ostara_CaseStudy_ThamesWater_web.pdf,

20. Oehmen, A., Lemos, P.C., Carvalho, G., Yuan, Z., Keller, J., Blackall, L.L., and Reis, M.A.M. Advances in Enhanced Biological Phosphorus Removal: From Micro to Macro Scale. *Water Research* **41**, 2271-2300 (2007).

21. Huang, W.E., Li, M., Jarvis, R.M., Goodacre, R., and Banwart, S.A., *Shining Light on the Microbial World: The Application of Raman Microspectroscopy*, in *Advances in Applied Microbiology, Vol 70*, A.I. Laskin, S. Sariaslani, and G.M. Gadd, Editors. 2010. p. 153-186.

22. Blackall, L.L., Crocetti, G., Saunders, A.M., and Bond, P.L. A Review and Update of the Microbiology of Enhanced Biological Phosphorus Removal in Wastewater Treatment Plants. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **81**, 681-691 (2002).

23. Barnard, J.L. Biological Nutrient Removal without Addition of Chemicals. *Water Research* **9**, 485-490 (1975).

24. Rao, N.N., Gomez-Garcia, M.R., and Kornberg, A., *Inorganic Polyphosphate: Essential for Growth and Survival*, in *Annual Review of Biochemistry*. 2009. p. 605-647.

25. Gebremariam, S.Y., Beutel, M.W., Christian, D., and Hess, T.F. Research Advances and Challenges in the Microbiology of Enhanced Biological Phosphorus Removal-a Critical Review. *Water Environment Research* **83**, 195-219 (2011).

26. Mino, T., Van Loosdrecht, M.C.M., and Heijnen, J.J. Microbiology and Biochemistry of the Enhanced Biological Phosphate Removal Process. *Water Research* **32**, 3193-3207 (1998).

27. Neethling, J., Blake, B., Benisch, M., Gu, A.Z., Stephens, H., Stensel, D., and Moore, R., *Factors Influencing the Reliability of Enhanced Biological Phosphorus Removal.* Water Environment Research Foundation 01-CTS-3

28. Forbes, C.M., O'Leary, N.D., Dobson, A.D., and Marchesi, J.R. The Contribution of 'Omic'-Based Approaches to the Study of Enhanced Biological Phosphorus Removal Microbiology. *FEMS Microbiology Ecology* **69**, 1-15 (2009).

29. Van Loosdrecht, M.C.M., Lopez-Vazquez, C.M., Meijer, S.C.F., Hooijmans, C.M., and Brdjanovic, D. Twenty-Five Years of ASM1: Past, Present and Future of Wastewater Treatment Modelling. *Journal of Hydroinformatics* **17**, 697-718 (2015).

30. Lopez-Vazquez, C.M., Oehmen, A., Hooijmans, C.M., Brdjanovic, D., Gijzen, H.J., Yuan, Z.G., and van Loosdrecht, M.C.M. Modeling the PAO-GAO Competition: Effects of Carbon Source, pH and Temperature. *Water Research* **43**, 450-462 (2009).

31. Filipe, C.D.M., Daigger, G.T., and Grady, C.P.L. pH as a Key Factor in the Competition between Glycogen-Accumulating Organisms and Phosphorus-Accumulating Organisms. *Water Environment Research* **73**, 223-232 (2001).

32. Basu, S., Fries, M., Majed, N., and Onnis-Hayden, A., *Enhanced Biological Phosphorus Removal*, in *Operation of Nutrient Removal Facilities WEF Manual of Practice No.37*. 2013, WEF: Alexandria, VA, USA.

33. Hesselmann, R.P.X., Werlen, C., Hahn, D., van der Meer, J.R., and Zehnder, A.J.B. Enrichment, Phylogenetic Analysis and Detection of a Bacterium That Performs Enhanced Biological Phosphate Removal in Activated Sludge. *Systematic and Applied Microbiology* **22**, 454-465 (1999).

34. He, S., Bishop, F.I., and McMahon, K.D. Bacterial Community and "Candidatus Accumulibacter" Population Dynamics in Laboratory-Scale Enhanced Biological Phosphorus Removal Reactors. *Applied and Environmental Microbiology* **76**, 5479-5487 (2010).

35. Kong, Y.H., Nielsen, J.L., and Nielsen, P.H. Microautoradiographic Study of Rhodocyclus-Related Polyphosphate Accumulating Bacteria in Full-Scale Enhanced Biological Phosphorus Removal Plants. *Applied and Environmental Microbiology* **70**, 5383-5390 (2004).

36. Seviour, R.J. and McIlroy, S. The Microbiology of Phosphorus Removal in Activated Sludge Processes - the Current State of Play. *Journal of Microbiology* **46**, 115-124 (2008).

37. Kong, Y.H., Nielsen, J.L., and Nielsen, P.H. Identity and Ecophysiology of Uncultured Actinobacterial Polyphosphate-Accumulating Organisms in Full-Scale Enhanced Biological Phosphorus Removal Plants. *Applied and Environmental Microbiology* **71**, 4076-4085 (2005).

38. Seviour, R.J., Kragelund, C., Kong, Y., Eales, K., Nielsen, J.L., and Nielsen, P.H. Ecophysiology of the Actinobacteria in Activated Sludge Systems. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **94**, 21-33 (2008).

39. He, S., Gall, D.L., and McMahon, K.D. "Candidatus Accumulibacter" Population Structure in Enhanced Biological Phosphorus Removal Sludges as Revealed by Polyphosphate Kinase Genes. *Applied and Environmental Microbiology* **73**, 5865-5874 (2007).

40. Carvalho, G., Lemos, P.C., Oehmen, A., and Reis, M.A.M. Denitrifying Phosphorus Removal: Linking the Process Performance with the Microbial Community Structure. *Water Research* **41**, 4383-4396 (2007).

41. Erdal, U.G., Erdal, Z.K., Daigger, G.T., and Randall, C.W. Is It PAO-GAO Competition or Metabolic Shift in EBPR System? Evidence from an Experimental Study. *Water Science and Technology* **58**, 1329-1334 (2008).

42. Yuan, Z., Pratt, S., and Batstone, D.J. Phosphorus Recovery from Wastewater through Microbial Processes. *Current Opinion in Biotechnology* **23**, 878-883 (2012).

43. Hirota, R., Kuroda, A., Kato, J., and Ohtake, H. Bacterial Phosphate Metabolism and Its Application to Phosphorus Recovery and Industrial Bioprocesses. *Journal of Bioscience and Bioengineering* **109**, 423-432 (2010).

44. Everall, N.J. Confocal Raman Microscopy: Common Errors and Artefacts. *Analyst* **135**, 2512-2522 (2010).

45. Puppels, G.J., Demul, F.F.M., Otto, C., Greve, J., Robertnicoud, M., Arndtjovin, D.J., and Jovin, T.M. Studying Single Living Cells and Chromosomes by Confocal Raman Microspectroscopy. *Nature* **347**, 301-303 (1990).

46. Pahlow, S., Meisel, S., Cialla-May, D., Weber, K., Rosch, P., and Popp, J. Isolation and Identification of Bacteria by Means of Raman Spectroscopy. *Advanced Drug Delivery Reviews* **89**, 105-120 (2015).

47. Li, M.Q., Xu, J., Romero-Gonzalez, M., Banwart, S.A., and Huang, W.E. Single Cell Raman Spectroscopy for Cell Sorting and Imaging. *Current Opinion in Biotechnology* **23**, 56-63 (2012).

48. Zumbusch, A., Holtom, G.R., and Xie, X.S. Three-Dimensional Vibrational Imaging by Coherent Anti-Stokes Raman Scattering. *Physical Review Letters* **82**, 4142-4145 (1999).

49. Tipping, W.J., Lee, M., Serrels, A., Brunton, V.G., and Hulme, A.N. Stimulated Raman Scattering Microscopy: An Emerging Tool for Drug Discovery. *Chemical Society Reviews* **45**, 2075-2089 (2016).

50. Schie, I.W., Krafft, C., and Popp, J. Applications of Coherent Raman Scattering Microscopies to Clinical and Biological Studies. *Analyst* **140**, 3897-3909 (2015).

51. Huh, Y.S., Chung, A.J., and Erickson, D. Surface Enhanced Raman Spectroscopy and Its Application to Molecular and Cellular Analysis. *Microfluidics and Nanofluidics* **6**, 285-297 (2009).

52. Willemse-Erix, D.F.M., Scholtes-Timmerman, M.J., Jachtenberg, J.-W., van Leeuwen, W.B., Horst-Kreft, D., Schut, T.C.B., Deurenberg, R.H., Puppels, G.J., van Belkum, A., Vos, M.C., and Maquelin, K. Optical Fingerprinting in Bacterial Epidemiology: Raman Spectroscopy as a Real-Time Typing Method. *Journal of Clinical Microbiology* **47**, 652-659 (2009).

53. Maquelin, K., Choo-Smith, L.P., Endtz, H.P., Bruining, H.A., and Puppels, G.J. Rapid Identification of Candida Species by Confocal Raman Micro Spectroscopy. *Journal of Clinical Microbiology* **40**, 594-600 (2002).

54. Maquelin, K., Kirschner, C., Choo-Smith, L.P., Ngo-Thi, N.A., van Vreeswijk, T., Stammler, M., Endtz, H.P., Bruining, H.A., Naumann, D., and Puppels, G.J. Prospective Study of the Performance of Vibrational Spectroscopies for Rapid Identification of Bacterial and Fungal Pathogens Recovered from Blood Cultures. *Journal of Clinical Microbiology* **41**, 324-329 (2003).

55. Almarashi, J.F.M., Kapel, N., Wilkinson, T.S., and Telle, H.H. Raman Spectroscopy of Bacterial Species and Strains Cultivated under Reproducible Conditions. *Spectroscopy-an International Journal* **27**, 361-365 (2012).

56. Buijtels, P.C.A.M., Willemse-Erix, H.F.M., Petit, P.L.C., Endtz, H.P., Puppels, G.J., Verbrugh, H.A., van Belkum, A., van Soolingen, D., and Maquelin, K. Rapid Identification of Mycobacteria by Raman Spectroscopy. *Journal of Clinical Microbiology* **46**, 961-965 (2008).

57. van de Vossenberg, J., Tervahauta, H., Maquelin, K., Blokker-Koopmans, C.H.W., Uytewaal-Aarts, M., van der Kooij, D., van Wezel, A.P., and van der Gaag, B. Identification of Bacteria in Drinking Water with Raman Spectroscopy. *Analytical Methods* **5**, 2679-2687 (2013).

58. Hutsebaut, D., Vandroemme, J., Heyrman, J., Dawyndt, P., Vandenabeele, P., Moens, L., and de vos, P. Raman Microspectroscopy as an Identification Tool within the Phylogenetically Homogeneous 'Bacillus Subtilis'-Group. *Systematic and Applied Microbiology* **29**, 650-660 (2006).

59. Ciobota, V., Burkhardt, E.-M., Schumacher, W., Roesch, P., Kuesel, K., and Popp, J. The Influence of Intracellular Storage Material on Bacterial Identification by Means of Raman Spectroscopy. *Analytical and Bioanalytical Chemistry* **397**, 2929-2937 (2010).

60. De Gelder, J., De Gussem, K., Vandenabeele, P., De Vos, P., and Moens, L. Methods for Extracting Biochemical Information from Bacterial Raman Spectra: An Explorative Study on Cupriavidus Metallidurans. *Analytica Chimica Acta* **585**, 234-240 (2007).

61. Beer, T., Tanaka, Z., Netzter, N., Rothschild, L.J., and Chen, B., *Analysis of Uncultured Extremophilic Snow Algae by Non-Invasive Single Cell Raman Spectroscopy*, in *Instruments, Methods, and Missions for Astrobiology XIV*, R.B. Hoover, Editor. 2011.

62. Watanabe, S., Yamanaka, M., Sakai, A., Sawada, K., and Iwasa, T. Laser Raman Spectroscopic Study on Magnetite Formation in Magnetotactic Bacteria. *Materials Transactions* **49**, 874-878 (2008).

63. Avetisyan, A., Jensen, J.B., and Huser, T. Monitoring Trehalose Uptake and Conversion by Single Bacteria Using Laser Tweezers Raman Spectroscopy. *Analytical Chemistry* **85**, 7264-7270 (2013).

64. Wu, H., Volponi, J.V., Oliver, A.E., Parikh, A.N., Simmons, B.A., and Singh, S. In Vivo Lipidomics Using Single-Cell Raman Spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3809-3814 (2011).

65. De Gelder, J., Willemse-Erix, D., Scholtes, M.J., Sanchez, J.I., Maquelin, K., Vandenabeele, P., De Boever, P., Puppels, G.J., Moens, L., and De Vos, P. Monitoring Poly(3-Hydroxybutyrate) Production in Cupriavidus Necator DSM 428 (H16) with Raman Spectroscopy. *Analytical Chemistry* **80**, 2155-2160 (2008).

66. Vitek, P., Camara-Gallego, B., Edwards, H.G.M., Jehlicka, J., Ascaso, C., and Wierzchos, J. Phototrophic Community in Gypsum Crust from the Atacama Desert Studied by Raman Spectroscopy and Microscopic Imaging. *Geomicrobiology Journal* **30**, 399-410 (2013).

67. Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., and Wagner, M. Raman-Fish: Combining Stable-Isotope Raman Spectroscopy and Fluorescence in Situ Hybridization for the Single Cell Analysis of Identity and Function. *Environmental Microbiology* **9**, 1878-1889 (2007).

68. Majed, N., Matthaus, C., Diem, M., and Gu, A.Z. Evaluation of Intracellular Polyphosphate Dynamics in Enhanced Biological Phosphorus Removal Process Using Raman Microscopy. *Environmental Science & Technology* **43**, 5436-5442 (2009).

69. Omelon, S., Georgiou, J., Variola, F., and Dean, M.N. Colocation and Role of Polyphosphates and Alkaline Phosphatase in Apatite Biomineralization of Elasmobranch Tesserae. *Acta Biomaterialia* **10**, 3899-3910 (2014).

70. Bednarova, L., Palacky, J., Bauerova, V., Hruskova-Heidingsfeldova, O., Pichova, I., and Mojzes, P. Raman Microspectroscopy of the Yeast Vacuoles. *Spectroscopy-an International Journal* **27**, 503-507 (2012).

71. Eder, S.H.K., Gigler, A.M., Hanzlik, M., and Winklhofer, M. Sub-Micrometer-Scale Mapping of Magnetite Crystals and Sulfur Globules in Magnetotactic Bacteria Using Confocal Raman Micro-Spectrometry. *PLOS One* **9**, (2014).

72. Majed, N. and Gu, A.Z. Application of Raman Microscopy for Simultaneous and Quantitative Evaluation of Multiple Intracellular Polymers Dynamics Functionally Relevant to Enhanced Biological Phosphorus Removal Processes. *Environmental Science & Technology* **44**, 8601-8608 (2010).

73. Choo-Smith, L.P., Maquelin, K., van Vreeswijk, T., Bruining, H.A., Puppels, G.J., Thi, N.A.G., Kirschner, C., Naumann, D., Ami, D., Villa, A.M., Orsini, F., Doglia, S.M., Lamfarraj, H., Sockalingum, G.D., Manfait, M., Allouch, P., and Endtz, H.P. Investigating Microbial (Micro)Colony Heterogeneity by Vibrational Spectroscopy. *Applied and Environmental Microbiology* **67**, 1461-1469 (2001).

74. Majed, N., Chernenko, T., Diem, M., and Gu, A.Z. Identification of Functionally Relevant Populations in Enhanced Biological Phosphorus Removal Processes Based on Intracellular Polymers Profiles and Insights into the Metabolic Diversity and Heterogeneity. *Environmental Science & Technology* **46**, 5010-5017 (2012).

75. Eaton, A.D., Clesceri, L.S., Greenberg, A.E., Franson, M.A.H., American Public Health Association., American Water Works Association., and Water Environment Federation., *Standard Methods for the Examination of Water and Wastewater*. 19th edition. ed. 1995, Washington, DC: Published jointly by: American Public Health Association, American Water Works Association, Water Environment Federation.

76. Schuler, A.J. and Jenkins, D. Enhanced Biological Phosphorus Removal from Wastewater by Biomass with Different Phosphorus Contents, Part I: Experimental Results and Comparison with Metabolic Models. *Water Environment Research* **75**, 485-498 (2003).

77. Bond, P.L., Erhart, R., Wagner, M., Keller, J., and Blackall, L.L. Identification of Some of the Major Groups of Bacteria in Efficient and Nonefficient Biological Phosphorus Removal Activated Sludge Systems. *Applied and Environmental Microbiology* **65**, 4077-4084 (1999).

78. Pijuan, M., Casas, C., and Baeza, J.A. Polyhydroxyalkanoate Synthesis Using Different Carbon Sources by Two Enhanced Biological Phosphorus Removal Microbial Communities. *Process Biochemistry* **44**, 97-105 (2009).

79. Carvalheira, M., Oehmen, A., Carvalho, G., and Reis, M.A.M. The Effect of Substrate Competition on the Metabolism of Polyphosphate Accumulating Organisms (PAOs). *Water Research* **64**, 149-159 (2014).

80. Whang, L.M. and Park, J.K. Competition between Polyphosphate- and Glycogen-Accumulating Organisms in Enhanced-Biological-Phosphorus-Removal Systems: Effect of Temperature and Sludge Age. *Water Environment Research* **78**, 4-11 (2006).

81. Met Office. *2015 Weather Summaries*. [Accessed: August 2016]; Available from: http://www.metoffice.gov.uk/climate/uk/summaries/2015.

82. Zhang, T., Liu, Y., and Fang, H.H.P. Effect of pH Change on the Performance and Microbial Community of Enhanced Biological Phosphate Removal Process. *Biotechnology and Bioengineering* **92**, 173-182 (2005).

83. Schuler, A.J. and Jenkins, D. Effects of pH on Enhanced Biological Phosphorus Removal Metabolisms. *Water Science and Technology* **46**, 171-178 (2002).

84. Brdjanovic, D., Slamet, A., van Loosdrecht, M.C.M., Hooijmans, C.M., Alaerts, G.J., and Heijnen, J.J. Impact of Excessive Aeration on Biological Phosphorus Removal from Wastewater. *Water Research* **32**, 200-208 (1998).

85. Griffiths, P.C., Stratton, H.M., and Seviour, R.J. Environmental Factors Contributing to the "G Bacteria" Population in Full-Scale EBPR Plants. *Water Science and Technology* **46**, 185-192 (2002).

86. *Commission of the European Communities. Directive Concerning Urban Wastewater Treatment (91/271/EEC). Official Journal, 1991:L135/40.*

87. Raman, C.V. and Krishnan, K.S. A New Type of Secondary Radiation. *Nature* **121**, 501-502 (1928).

88. Keresztury, G., *Raman Spectroscopy: Theory* in *Handbook of Vibrational Spectroscopy. Volume 1: Theory and Instrumentation*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, John Wiley and Sons Ltd. : Chichester, UK.

89. Turrell, G., *The Raman Effect*, in *Raman Microscopy Developments and Applications*, G. Turrell and J. Corset, Editors. 1996, Elesevier Ltd.

90. Larkin, P., *Infrared and Raman Spectroscopy Principles and Spectral Interpretation*. 2011, Amsterdam ; Boston: Elsevier,.

91. Opilik, L., Schmid, T., and Zenobi, R., *Modern Raman Imaging: Vibrational Spectroscopy on the Micrometer and Nanometer Scales*, in *Annual Review of Analytical Chemistry, Vol 6*, R.G. Cooks and J.E. Pemberton, Editors. 2013. p. 379-398.

92. Tobias, R.S. Raman Spectroscopy in Inorganic Chemistry I. Theory. *Journal of Chemical Education* **44**, 2-8 (1967).

93. Mukerji, I. *What Is Raman Spectroscopy*. About Sickle Cell Disease [Accessed: December 2015]; Available from: http://www.sicklecellinfo.net/raman.htm

94. Hollas, J.M. *Basic Atomic and Molecular Spectroscopy*. 2002, Cambridge, UK: The Royal Society of Chemistry.

95. Davidson, B.K., *A Raman Spectroscopic Investigation of the Murine Oocyte*. 2010, University of Edinburgh.

96. Bradley, M.S. Lineshapes in IR and Raman Spectroscopy: A Primer. *Spectroscopy* **30**, 42-46 (2015).

97. Izumi, C.M.S. and Temperini, M.L.A. FT-Raman Investigation of Biodegradable Polymers: Poly(3-Hydroxybutyrate) and Poly(3-Hydroxybutyrate-Co-3-Hydroxyvalerate). *Vibrational Spectroscopy* **54**, 127-132 (2010).

98. Batchelder, D.N. (1993). *Confocal Spectroscopy*. EP 0542962 A1.

99. Williams, K.P.J., Pitt, G.D., Batchelder, D.N., and Kip, B.J. Confocal Raman Microspectroscopy Using a Stigmatic Spectrograph and CCD Detector. *Applied Spectroscopy* **48**, 232-235 (1994).

100. Hodkiewicz, J. and Deck, F., *Technical Note 52043: Evaluating Spectral Resolution on a Raman Spectrometer*. Thermo Scientific, Madison, WI, USA,

101. Inoué, S., *Foundations of Confocal Scanned Imaging in Light Microscopy*, in *Handbook of Biological Confocal Microscopy*, J.B. Pawley, Editor. 2006, Springer US: Boston, MA: Boston, MA.

102. Horiba Scientific *What Is the Spatial Resolution of a Raman Microscope?* [Accessed: December 2015]; Available from: http://www.horiba.com/scientific/products/raman-spectroscopy/raman-academy/raman-faqs/what-is-the-spatial-resolution-of-a-raman-microscope/.

103. Matthaeus, C., Bird, B., Miljkovic, M., Chernenko, T., Romeo, M., and Diem, M., *Infrared and Raman Microscopy in Cell Biology*, in *Biophysical Tools for Biologists, Vol 2: In Vivo Techniques*, J.J. Correia and H.W. Detrich, Editors. 2008. p. 275-308.

104. OriginLab.*Gaussamp Fit Function*. [Accessed: December 2015]; Available from: http://www.originlab.com/doc/Origin-Help/GaussAmp-FitFunc.

105. Adar, F. Depth Resolution of the Raman Microscope: Optical Limitations and Sample Characteristics. *Spectroscopy* **25**, 16-23 (2010).

106. Lee, E., *Imaging Modes*, in *Raman Imaging: Techniques and Applications*, A. Zoubir, Editor., Springer-Verlag: Berlin p. 10-12.

107. R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [Accessed: December 2015]; Available from: https://www.R-project.org/.

108. de Juan, A. and Tauler, R. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Critical Reviews in Analytical Chemistry* **36**, 163-176 (2006).

109. Jiang, J.H., Liang, Y.Z., and Ozaki, Y. Principles and Methodologies in Self-Modeling Curve Resolution. *Chemometrics and Intelligent Laboratory Systems* **71**, 1-12 (2004).

110. Barnes, R.J., Dhanoa, M.S., and Lister, S.J. Standard Normal Variate Transformation and De-Trending of near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy* **43**, 772-777 (1989).

111. Martens, H., *Multivariate Calibration*, ed. T. Næs. 1991, Chichester : Wiley.

112. Martens, H., Nielsen, J.P., and Engelsen, S.B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to near-Infrared Transmission Analysis of Powder Mixtures. *Analytical Chemistry* **75**, 394-404 (2003).

113. Palacky, J., Mojzes, P., and Bok, J. SVD-Based Method for Intensity Normalization, Background Correction and Solvent Subtraction in Raman Spectroscopy Exploiting the Properties of Water Stretching Vibrations. *Journal of Raman Spectroscopy* **42**, 1528-1539 (2011).

114. Beattie, J.R. and McGarvey, J.J. Estimation of Signal Backgrounds on Multivariate Loadings Improves Model Generation in Face of Complex Variation in Backgrounds and Constituents. *Journal of Raman Spectroscopy* **44**, 329-338 (2013).

115. Schulze, G., Jirasek, A., Yu, M.M.L., Lim, A., Turner, R.F.B., and Blades, M.W. Investigation of Selected Baseline Removal Techniques as Candidates for Automated Implementation. *Applied Spectroscopy* **59**, 545-574 (2005).

116. Westerhuis, J.A., de Jong, S., and Smilde, A.K. Direct Orthogonal Signal Correction. *Chemometrics and Intelligent Laboratory Systems* **56**, 13-25 (2001).

117. Friedrichs, M.S. A Model-Free Algorithm for the Removal of Base-Line Artifacts. *Journal of Biomolecular NMR* **5**, 147-153 (1995).

118. Mark, H. and Workman, J. Derivatives in Spectroscopy - Part III - Computing the Derivative. *Spectroscopy* **18**, 106-111 (2003).

119. Brown, C.D., Vega-Montoto, L., and Wentzell, P.D. Derivative Preprocessing and Optimal Corrections for Baseline Drift in Multivariate Calibration. *Applied Spectroscopy* **54**, 1055-1068 (2000).

120. Mosierboss, P.A., Lieberman, S.H., and Newbery, R. Fluorescence Rejection in Raman-Spectroscopy by Shifted-Spectra, Edge-Detection, and FFT Filtering Techniques. *Applied Spectroscopy* **49**, 630-638 (1995).

121. Tan, H.W. and Brown, S.D. Wavelet Analysis Applied to Removing Non-Constant, Varying Spectroscopic Background in Multivariate Calibration. *Journal of Chemometrics* **16**, 228-240 (2002).

122. Kneen, M.A. and Annegarn, H.J. Algorithm for Fitting XRF, SEM and PIXE X-Ray Spectra Backgrounds. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms* **109**, 209-213 (1996).

123. Schulze, H.G., Foist, R.B., Okuda, K., Ivanov, A., and Turner, R.F.B. A Small-Window Moving Average-Based Fully Automated Baseline Estimation Method for Raman Spectra. *Applied Spectroscopy* **66**, 757-764 (2012).

124. Jirasek, A., Schulze, G., Yu, M.M.L., Blades, M.W., and Turner, R.F.B. Accuracy and Precision of Manual Baseline Determination. *Applied Spectroscopy* **58**, 1488-1499 (2004).

125. Dietrich, W., Rüdel, C.H., and Neumann, M. Fast and Precise Automatic Baseline Correction of One- and Two-Dimensional NMR Spectra. *Journal of Magnetic Resonance (1969)* **91**, 1-11 (1991).

126. Coombes, K.R., Fritsche, H.A., Clarke, C., Chen, J.N., Baggerly, K.A., Morris, J.S., Xiao, L.C., Hung, M.C., and Kuerer, H.M. Quality Control and Peak Finding for Proteomics Data Collected from Nipple Aspirate Fluid by Surface-Enhanced Laser Desorption and Ionization. *Clinical Chemistry* **49**, 1615-1623 (2003).

127. Lieber, C.A. and Mahadevan-Jansen, A. Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Applied Spectroscopy* **57**, 1363-1367 (2003).

128. Zhao, J., Lui, H., McLean, D.I., and Zeng, H. Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy. *Applied Spectroscopy* **61**, 1225-1232 (2007).

129. Cao, A., Pandya, A.K., Serhatkulu, G.K., Weber, R.E., Dai, H., Thakur, J.S., Naik, V.M., Naik, R., Auner, G.W., Rabah, R., and Freeman, D.C. A Robust Method for Automated Background Subtraction of Tissue Fluorescence. *Journal of Raman Spectroscopy* **38**, 1199-1205 (2007).

130. Beier, B.D. and Berger, A.J. Method for Automated Background Subtraction from Raman Spectra Containing Known Contaminants. *Analyst* **134**, 1198-1202 (2009).

131. Liu, J., Sun, J., Huang, X., Li, G., and Liu, B. Goldindec: A Novel Algorithm for Raman Spectrum Baseline Correction. *Applied Spectroscopy* **69**, 834-842 (2015).

132. Eilers, P.H.C. Parametric Time Warping. *Analytical Chemistry* **76**, 404-411 (2004).

133. Zhang, Z.-M., Chen, S., and Liang, Y.-Z. Baseline Correction Using Adaptive Iteratively Reweighted Penalized Least Squares. *Analyst* **135**, 1138-1146 (2010).

134. Cadusch, P.J., Hlaing, M.M., Wade, S.A., McArthur, S.L., and Stoddart, P.R. Improved Methods for Fluorescence Background Subtraction from Raman Spectra. *Journal of Raman Spectroscopy* **44**, 1587-1595 (2013).

135. Ruckstuhl, A.F., Jacobson, M.P., Field, R.W., and Dodd, J.A. Baseline Subtraction Using Robust Local Regression Estimation. *Journal of Quantitative Spectroscopy & Radiative Transfer* **68**, 179-193 (2001).

136. Afseth, N.K., Segtnan, V.H., and Wold, J.P. Raman Spectra of Biological Samples: A Study of Preprocessing Methods. *Applied Spectroscopy* **60**, 1358-1367 (2006).

137. Heraud, P., Wood, B.R., Beardall, J., and McNaughton, D. Effects of Pre-Processing of Raman Spectra on in Vivo Classification of Nutrient Status of Microalgal Cells. *Journal of Chemometrics* **20**, 193-197 (2006).

138. Leger, M.N. and Ryder, A.G. Comparison of Derivative Preprocessing and Automated Polynomial Baseline Correction Method for Classification and

Quantification of Narcotics in Solid Mixtures. *Applied Spectroscopy* **60**, 182-193 (2006).

139. Liland, K.H., Almoy, T., and Mevik, B.-H. Optimal Choice of Baseline Correction for Multivariate Calibration of Spectra. *Applied Spectroscopy* **64**, 1007-1016 (2010).

140. Liland, K.H. and Mevik, B.-H.(2015) *Baseline Correction of Spectra. R Package Version 1.2-1*. [Accessed: December 2015]; Available from: http://CRAN.R-project.org/package=baseline.

141. Maquelin, K., Choo-Smith, L.P., van Vreeswijk, T., Endtz, H.P., Smith, B., Bennett, R., Bruining, H.A., and Puppels, G.J. Raman Spectroscopic Method for Identification of Clinically Relevant Microorganisms Growing on Solid Culture Medium. *Analytical Chemistry* **72**, 12-19 (2000).

142. Petersen, P., *Linear Algebra*. 2012, New York, N.Y.: New York, N.Y. : Springer.

143. Berger, A.J.*Biomedical Spectroscopy Lab, University of Rochester: Spectrum-Based Method for Iterative Removal of Fluorescence*. [Accessed: December 2015]; Available from: https://sites.google.com/site/bslrochester/downloads-1

144. Eilers, P.H.C. A Perfect Smoother. *Analytical Chemistry* **75**, 3631-3636 (2003).

145. Whittaker, E.T. On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society* **41**, 63-75 (1923).

146. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 829-836 (1979).

147. Kornberg, A., Rao, N.N., and Ault-Riche, D. Inorganic Polyphosphate: A Molecule of Many Functions. *Annual Review of Biochemistry* **68**, 89-125 (1999).

148. Peoples, O.P. and Sinskey, A.J. Poly-Beta-Hydroxybutyrate (Phb) Biosynthesis in Alcaligenes-Eutrophus H16 - Identification and Characterization of the PHB Polymerase Gene (PHBC). *Journal of Biological Chemistry* **264**, 15298-15303 (1989).

149. Häggström, M., *Medical Gallery of Mikael Häggström 2014*. 2014: Wikiversity Journal of Medicine 1 (2).

150. Maquelin, K., Choo-Smith, L.P., Kirschner, C., Ngo-Thi, N.A., Naumann, D., and Puppels, G.J., *Vibrational Spectroscopic Studies of Microorganisms*, in *Handbook of Vibrational Spectroscopy*. 2006, John Wiley & Sons, Ltd.

151. Notingher, I. and Hench, L.L. Raman Microspectroscopy: A Noninvasive Tool for Studies of Individual Living Cells in Vitro. *Expert Review of Medical Devices* **3**, 215-234 (2006).

152. Miyazawa, T., Shimanouchi, T., and Mizushima, S.I. Normal Vibrations of N-Methylacetamide. *Journal of Chemical Physics* **29**, 611-616 (1958).

153. Movasaghi, Z., Rehman, S., and Rehman, I.U. Raman Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews* **42**, 493-541 (2007).

154. Kulakova, A.N., Hobbs, D., Smithen, M., Pavlov, E., Gilbert, J.A., Quinn, J.P., and McGrath, J.W. Direct Quantification of Inorganic Polyphosphate in Microbial Cells Using 4 '-6-Diamidino-2-Phenylindole (DAPI). *Environmental Science & Technology* **45**, 7799-7803 (2011).

155. Diaz, J.M. and Ingall, E.D. Fluorometric Quantification of Natural Inorganic Polyphosphate. *Environmental Science & Technology* **44**, 4665-4671 (2010).

156. Ray, N.H. Oxide Glasses of Very Low Softening Point .3. Study of Potassium Lead Phosphate Glasses by Raman-Spectroscopy. *Glass Technology* **16**, 107-108 (1975).

157. De Jager, H.-J. and Heyns, A.M. Study of the Hydrolysis of Sodium Polyphosphate in Water Using Raman Spectroscopy. *Applied Spectroscopy* **52**, 808-814 (1998).

158. Koller, M. and Rodríguez-Contreras, A. Techniques for Tracing PHA-Producing Organisms and for Qualitative and Quantitative Analysis of Intra- and Extracellular PHA. *Engineering in Life Sciences* **15**, 558-581 (2015).

159. Oehmen, A., Yuan, Z.G., Blackall, L.L., and Keller, J. Comparison of Acetate and Propionate Uptake by Polyphosphate Accumulating Organisms and Glycogen Accumulating Organisms. *Biotechnology and Bioengineering* **91**, 162-168 (2005).

160. Oehmen, A., Yuan, Z., Blackall, L.L., and Keller, J. Short-Term Effects of Carbon Source on the Competition of Polyphosphate Accumulating Organisms and Glycogen Accumulating Organisms. *Water Science and Technology* **50**, 139-144 (2004).

161. Gałat, A. Study of the Raman Scattering and Infrared Absorption Spectra of Branched Polysaccharides. *Acta biochimica Polonica* **27**, 135-142 (1980).

162. Konorov, S.O., Schulze, H.G., Atkins, C.G., Piret, J.M., Aparicio, S.A., Turner, R.F.B., and Blades, M.W. Absolute Quantification of Intracellular Glycogen Content in Human Embryonic Stem Cells with Raman Microspectroscopy. *Analytical Chemistry* **83**, 6254-6258 (2011).

163. Parker, F.S., *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*. 1983, New York and London: Plenum Press. 337-343.

164. Kim, J.M., Lee, H.J., Kim, S.Y., Song, J.J., Park, W., and Jeon, C.O. Analysis of the Fine-Scale Population Structure of "Candidatus Accumulibacter Phosphatis" in Enhanced Biological Phosphorus Removal Sludge, Using Fluorescence in Situ Hybridization and Flow Cytometric Sorting. *Applied and Environmental Microbiology* **76**, 3825-3835 (2010).

165. Zhang, Q., Zhang, P., Gou, H., Mou, C., Huang, W.E., Yang, M., Xu, J., and Ma, B. Towards High-Throughput Microfluidic Raman-Activated Cell Sorting. *Analyst* **140**, 6163-6174 (2015).

166. Sureka, K., Dey, S., Datta, P., Singh, A.K., Dasgupta, A., Rodrigue, S., Basu, J., and Kundu, M. Polyphosphate Kinase is Involved in Stress-Induced mprAB-SigE-rel Signalling in Mycobacteria. *Molecular Microbiology* **65**, 261-276 (2007).

167. Brown, M.R.W. and Kornberg, A. The Long and Short of It - Polyphosphate, ppk and Bacterial Survival. *Trends in Biochemical Sciences* **33**, 284-290 (2008).

168. Kiyono, M., Omura, H., Omura, T., Murata, S., and Pan-Hou, H. Removal of Inorganic and Organic Mercurials by Immobilized Bacteria Having mer-ppk Fusion Plasmids. *Applied Microbiology and Biotechnology* **62**, 274-278 (2003).

169. Grillo-Puertas, M., Ariane Schurig-Briccio, L., Rodriguez-Montelongo, L., Regina Rintoul, M., and Andrea Rapisarda, V. Copper Tolerance Mediated by Polyphosphate Degradation and Low-Affinity Inorganic Phosphate Transport System in Escherichia Coli. *BMC Microbiology* **14**, (2014).

170. Navarro, C.A., von Bernath, D., and Jerez, C.A. Heavy Metal Resistance Strategies of Acidophilic Bacteria and Their Acquisition: Importance for Biomining and Bioremediation. *Biological Research* **46**, 363-371 (2013).

171. Remonsellez, F., Orell, A., and Jerez, C.A. Copper Tolerance of the Thermoacidophilic Archaeon Sulfolobus Metallicus: Possible Role of Polyphosphate Metabolism. *Microbiology-SGM* **152**, 59-66 (2006).

172. Suzuki, Y. and Banfield, J.F. Resistance to, and Accumulation of, Uranium by Bacteria from a Uranium-Contaminated Site. *Geomicrobiology Journal* **21**, 113-121 (2004).

173. Serafim, L.S., Lemos, P.C., Levantesi, C., Tandoi, V., Santos, H., and Reis, M.A.M. Methods for Detection and Visualization of Intracellular Polymers Stored by Polyphosphate-Accumulating Microorganisms. *Journal of Microbiological Methods* **51**, 1-18 (2002).

174. Brown, M.R.W. and Kornberg, A. Inorganic Polyphosphate in the Origin and Survival of Species. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16085-16087 (2004).

175. Wood, H.G. and Clark, J.E. Biological Aspects of Inorganic Polyphosphates. *Annual Review of Biochemistry* **57**, 235-260 (1988).

176. Jing, S.R., Benefield, L.D., and Hill, W.E. Observations Relating to Enhanced Phosphorus Removal in Biological-Systems. *Water Research* **26**, 213-223 (1992).

177. Kulaev, I.S., Vagabov, V.M., and Kulakovskaya, T.V., *The Biochemistry of Inorganc Polyphosphates, 2nd Edition*. 2004, Chichester, UK: John Wiley & Sons, Ltd.

178. Nagel, L. Volutin. *Botanical Review* **14**, 174-184 (1948).

179. Müssig-Zufika, M., Kornmüller, A., Merkelbach, B., and Jekel, M. Isolation and Analysis of Intact Polyphosphate Chains from Activated Sludges Associated with Biological Phosphate Removal. *Water Research* **28**, 1725-1733 (1994).

180. Eixler, S., Selig, U., and Karsten, U. Extraction and Detection Methods for Polyphosphate Storage in Autotrophic Planktonic Organisms. *Hydrobiologia* **533**, 135-143 (2005).

181. Hupfer, M. and Gachter, R. Polyphosphate in Lake-Sediments - P-31 NMR-Spectroscopy as a Tool for its Identification. *Limnology and Oceanography* **40**, 610-617 (1995).

182. Damle, S.P. and Krishnan, P.S. Studies on the Role of Metaphosphate in Molds .1. Quantitative Studies on the Metachromatic Effect of Metaphosphate. *Archives of Biochemistry and Biophysics* **49**, 58-70 (1954).

183. Leitao, J.M., Lorenz, B., Bachinski, N., Wilhelm, C., Muller, W.E.G., and Schroder, H.C. Osmotic-Stress-Induced Synthesis and Degradation of Inorganic Polyphosphates in the Alga Phaeodactylum-Tricornutum. *Marine Ecology Progress Series* **121**, 279-288 (1995).

184. Lorenz, B., Munkner, J., Oliveira, M.P., Leitao, J.M., Muller, W.E.G., and Schroder, H.C. A Novel Method for Determination of Inorganic Polyphosphates Using the Fluorescent Dye Fura-2. *Analytical Biochemistry* **246**, 176-184 (1997).

185. Ohtomo, R., Sekiguchi, Y., Mimura, T., Saito, M., and Ezawa, T. Quantification of Polyphosphate: Different Sensitivities to Short-Chain Polyphosphate Using Enzymatic and Colorimetric Methods as Revealed by Ion Chromatography. *Analytical Biochemistry* **328**, 139-146 (2004).

186. Klauth, P., Pallerla, S.R., Vidaurre, D., Ralfs, C., Wendisch, V.F., and Schoberth, S.M. Determination of Soluble and Granular Inorganic Polyphosphate in Corynebacterium Glutamicum. *Applied Microbiology and Biotechnology* **72**, 1099-1106 (2006).

187. Mesquita, D.P., Amaral, A.L., Leal, C., Carvalheira, M., Cunha, J.R., Oehmen, A., Reis, M.A.M., and Ferreira, E.C. Monitoring Intracellular Polyphosphate Accumulation in Enhanced Biological Phosphorus Removal Systems by Quantitative Image Analysis. *Water Science and Technology* **69**, 2315-2323 (2014).

188. Aschar-Sobbi, R., Abramov, A.Y., Diao, C., Kargacin, M.E., Kargacin, G.J., French, R.J., and Pavlov, E. High Sensitivity, Quantitative Measurements of Polyphosphate Using a New DAPI-Based Approach. *Journal of Fluorescence* **18**, 859-866 (2008).

189. Ault-Riche, D., Fraley, C.D., Tzeng, C.M., and Kornberg, A. Novel Assay Reveals Multiple Pathways Regulating Stress-Induced Accumulations of Inorganic Polyphosphate in Escherichia Coli. *Journal of Bacteriology* **180**, 1841-1847 (1998).

190. Ohtomo, R., Sekiguchi, Y., Kojima, T., and Saito, M. Different Chain Length Specificity among Three Polyphosphate Quantification Methods. *Analytical Biochemistry* **383**, 210-216 (2008).

191. Amado, L. and Kuzminov, A. Polyphosphate Accumulation in Escherichia Coli in Response to Defects in DNA Metabolism. *Journal of Bacteriology* **191**, 7410-7416 (2009).

192. Clark, J.E. and Wood, H.G. Preparation of Standards and Determination of Sizes of Long-Chain Polyphosphates by Gel-Electrophoresis. *Analytical Biochemistry* **161**, 280-290 (1987).

193. Lee, A. and Whitesides, G.M. Analysis of Inorganic Polyphosphates by Capillary Gel Electrophoresis. *Analytical Chemistry* **82**, 6838-6846 (2010).

194. Cade-Menun, B.J. Characterizing Phosphorus in Environmental and Agricultural Samples by P-31 Nuclear Magnetic Resonance Spectroscopy. *Talanta* **66**, 359-371 (2005).

195. Hill, W.E., Benefield, L.D., and Jing, S.R. P-31 NMR-Spectroscopy Characterization of Polyphosphates in Activated-Sludge Exhibiting Enhanced Phosphorus Removal. *Water Research* **23**, 1177-1181 (1989).

196. Lambert, C., Weuster-Botz, D., Weichenhain, R., Kreutz, E.W., De Graaf, A.A., and Schoberth, S.M. Monitoring of Inorganic Polyphosphate Dynamics in Corynebacterium Glutamicum Using a Novel Oxygen Sparger for Real Time P-31 in Vivo Nmr. *Acta Biotechnologica* **22**, 245-260 (2002).

197. Lawrence, B.A., Suarez, C., Depina, A., Click, E., Kolodny, N.H., and Allen, M.M. Two Internal Pools of Soluble Polyphosphate in the Cyanobacterium Synechocystis sp. Strain PCC 6308: An in Vivo 31P NMR Spectroscopy Study. *Archives of Microbiology* **169**, 195-200 (1998).

198. Florentz, M., Granger, P., and Hartemann, P. Use of P-31 Nuclear Magnetic-Resonance Spectroscopy and Electron-Microscopy to Study Phosphorus-Metabolism of Microorganisms from Wastewaters. *Applied and Environmental Microbiology* **47**, 519-525 (1984).

199. Pereira, H., Lemos, P.C., Reis, M.A.M., Crespo, J., Carrondo, M.J.T., and Santos, H. Model for Carbon Metabolism in Biological Phosphorus Removal Processes Based on in Vivo C-13-NMR Labelling Experiments. *Water Research* **30**, 2128-2138 (1996).

200. Santos, M.M., Lemos, P.C., Reis, M.A.M., and Santos, H. Glucose Metabolism and Kinetics of Phosphorus Removal by the Fermentative Bacterium Microlunatus Phosphovorus. *Applied and Environmental Microbiology* **65**, 3920-3928 (1999).

201. Khoshmanesh, A., Cook, P.L.M., and Wood, B.R. Quantitative Determination of Polyphosphate in Sediments Using Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) Spectroscopy and Partial Least Squares Regression. *Analyst* **137**, 3704-3709 (2012).

202. Pelletier, M.J. Quantitative Analysis Using Raman Spectrometry. *Applied Spectroscopy* **57**, 20A-42A (2003).

203. Giles, J.H., Gilmore, D.A., and Denton, M.B. Quantitative Analysis Using Raman Spectroscopy without Spectral Standardization. *Journal of Raman Spectroscopy* **30**, 767-771 (1999).

204. Ab-Rahman, M.S. and Shuhaimi, N.I. The Effect of Temperature on the Performance of Uncooled Semiconductor Laser Diode in Optical Network. *Journal of Computer Science* **8**, 84-88 (2012).

205. Toptica Photonics. *Xtra II Single-Frequency High Power Diode Laser*. [Accessed: July 2016]; Available from: http://www.toptica.com/fileadmin/user_upload/Brochures/toptica_BR_XTRA_II.pdf.

206. Beattie, R.J., Bell, S.J., Farmer, L.J., Moss, B.W., and Desmond, P.D. Preliminary Investigation of the Application of Raman Spectroscopy to the Prediction of the Sensory Quality of Beef Silverside. *Meat Science* **66**, 903-913 (2004).

207. Bro, R. Multivariate Calibration: What Is in Chemometrics for the Analytical Chemist? *Analytica Chimica Acta* **500**, 185-194 (2003).

208. Vandenabeele, P., Jehlicka, J., Vitek, P., and Edwards, H.G.M. On the Definition of Raman Spectroscopic Detection Limits for the Analysis of Biomarkers in Solid Matrices. *Planetary and Space Science* **62**, 48-54 (2012).

209. Gu, A.Z., Saunders, A., Neethling, J.B., Stensel, H.D., and Blackall, L.L. Functionally Relevant Microorganisms to Enhanced Biological Phosphorus Removal Performance at Full-Scale Wastewater Treatment Plants in the United States. *Water Environment Research* **80**, 688-698 (2008).

210. Gaudy, A.F. and Blachly, T.R. A Study of the Biodegradability of Residual Cod. *Journal Water Pollution Control Federation* **57**, 332-338 (1985).

211. ICH, *Q2 (R1). Validation of Analytical Procedures: Text and Methodology.* ICH Harmonized Tripartite Guideline, International Conference on the Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Chicago, USA,November 2005.